Aspects of DNA Replication Initiator Proteins of *Escherichia coli* and its Bacteriophage

Christopher Anthony Love

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University

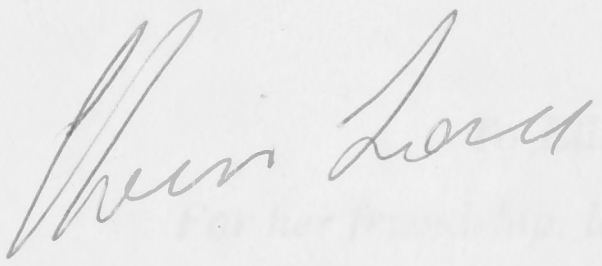
RESEARCH SCHOOL OF CHEMISTRY

THE AUSTRALIAN NATIONAL UNIVERSITY

October 1996

DECLARATION

The work described herein is the authors own work, unless otherwise stated, and was carried out within the Research School of Chemistry, The Australian National University, between May 1992 and October 1996. The research project required the application of various specialised techniques and hence the assistance of several people. The ultracentrifugation studies of DnaA were performed by Dr Peter Jeffrey at the John Curtin School of Medical Research, The Australian National University while electrospray ionisation mass spectrometry was carried out by Dr Greg Kilby. The contents of this thesis have not been included in any other work submitted by the author for another degree.



Christopher Love

To Ellen

For her friendship, love and support.

ACKNOWLEDGMENTS

I am in debt to my supervisor, Dr Nicholas Dixon, whom I would like to thank for his endless enthusiasm, support and motivation throughout my postgraduate research. His passion for science and friendship socially have contributed to making my post graduate course an enjoyable experience. I would also like to thank my advisors, Dr Wilfred Armarego and Dr David Ollis, for their support.

Special thanks to Penny Lilley, for all the excellent assistance, and Caroline Miles and Andrew Kralicek for proof reading my thesis. Thanks to John Barton Jeffrey Crowther, Nadá Dammerová, Blair Henderson, Thomas Huber, Shadi Moghaddas, Cameron Neylon, Chris Pennington, Peter Suffolk, Subash Vasudevan, Jigeon Yang for their friendship. I would also like to thank all the members of the Ollis and Torda groups.

I would especially like to thank Dr Greg Kilby for his assistance with the electrospray mass spectrometry and Dr Peter Jeffrey for performing the sedimentation equilibrium experiments. I must thank Karen Edwards and Dr Paul Carr for their advice on protein crystallography.

Thanks to Dr Andrew Torda, the only guy I know who is "bright and cheery" 24 hours a day. Thanks to my pool partner, Kirstie MacPherson, the "Dream Team" and all those other pool sharks.

To my family, especially my mother Beverley Antrobus, thanks for your love, encouragement and support. Without you, I would not be here.

PUBLICATIONS

JOURNAL ARTICLES

Love, C.A., Lilley, P.E. and Dixon, N.E. (1996) Stable high-copy-number bacteriophage λ promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* (in press)

Lilley, P.E., **Love, C.A.**, Argall, M.A., Elvin, C.E. and Dixon, N.E. Cloning and overproduction of the *Escherichia coli* DnaA replication initiator protein. (in preparation)

Love, C.A., Argall, M.A., Lilley, P.E. and Dixon, N.E. Purification and characterisation of the *Escherichia coli* DnaA replication initiator protein. (in preparation)

Love, C.A., Miles, C.S., Lilley, P.E. and Dixon, N.E. The effect of different nucleotide spacing between the ribosome binding site and start codon on levels of overproduction of the *Escherichia coli* DnaA and Tus proteins. (in preparation)

CONFERENCE ABSTRACTS

Love, C.A., Argall, M.E., Lilley, P.E. and Dixon, N.E. (1993) Large-scale purification of the *Escherichia coli* DnaA replication initiator protein. *Proc. Aust. Soc. Biochem. Mol. Biol.*

Love, C.A. and Dixon, N.E. (1995) Cloning and overexpression of bacteriophage rolling-circle initiator proteins. *Proc. Aust. Soc. Biochem. Mol. Biol.*

Dixon, N.E., Edwards, K.J., Henderson, B.R., Kralicek, A.V., Lilley, P.E., **Love, C.A.**, Miles, C.S., Neylon, D.C. and Yang, J.Y. (1996) Development of a method for *in vitro* directed molecular evolution of protein structure and function. Proc. East Coast Protein Meeting.

Henderson, B.R., Dixon, N.E., Edwards, K.J., Kralicek, A.V., Lilley, P.E., **Love, C.A.**, Miles, C.S., Neylon, D.C. and Yang, J.Y. (1996) Development of a method for *in vitro* directed molecular evolution of protein structure and function. Proc. Queenstown Molecular Biology Meeting.

Kralicek, A.V., Edwards, K.J., Henderson, B.R., Lilley, P.E., **Love, C.A.**, Miles, C.S., Neylon, D.C., Yang, J.Y. and Dixon, N.E. (1996) In vitro directed molecular evolution of protein function. *Proc. Aust. Soc. Biochem. Mol. Biol.*

ABBREVIATIONS

Aλ	absorbance measured at wavelength λ (nm) in cuvettes of 1-cm pathway
ADP	adenosine 5´-diphosphate
ATP	adenosine 5´-triphosphate
bp	base pair(s)
BSA	bovine serum albumin
Da	dalton(s)
Dam	deoxyadenosine methyltransferase
dATP	deoxyadenosine 5´-triphosphate
dCTP	deoxycytidine 5´-triphosphate
dGTP	deoxyguanosine 5´-triphosphate
dTTP	deoxythymidine 5´-triphosphate
dNTP	deoxynucleoside 5´-triphosphate
DnaA	the DnaA protein
dnaA	the gene encoding DnaA protein
DnaB	the DnaB protein
DnaC	the DnaC protein
ds	double strand(ed) (DNA)
DTT	dithiothreitol
E. coli	*Escherichia coli*
ESI-MS	electrospray ionisation mass spectometry
HEPES	4-(2-hydrxyethyl)-1-piperazine-ethanesulfonic acid
IHF	integration host factor
LB	Luria-Bertani medium
MOPS	3-(N-morpholino)propanesulfonic acid
PAGE	polyacrylamide gel electrophoresis
pas	primosome assembly site
p_R, p_L	major leftward and rightward promoters of phage λ, respectively
RBS	ribosome binding site
RF	replicative form
SDS	sodium dodecyl sulfate
ss	single strand(ed) (DNA)
SSB	single-stranded DNA binding protein
TCA	trichloroacetic acid
TEMED	NNN'N'-tetraethylenediamine
Tris	tris(hydroxymethyl)-aminomethane
ts	temperature-sensitive
Tus	termination utililisation substance

ABSTRACT

The genetic material of all organisms must be faithfully replicated once (and only once) per cell cycle. Chromosomal replication is controlled at the level of initiation, primarily by the action of the DnaA replication initiator protein at the origin of replication. The process of initiation of DNA replication of the *Escherichia coli* chromosome involves several proteins (DnaA, DnaB, DnaC and others) which work in a concerted fashion to open the duplex DNA and load the replication machinery. The *E. coli* bacteriophages M13 and ϕX174 rely on their host for the supply of all replication machinery with the exception of the phage-encoded replicator initiator proteins. These (Gene II of M13 and Gene A of ϕX174) are endonucleases which initiate rolling-circle replication in a manner completely different to that of DnaA in initiation of chromosomal replication.

Studies on three replication initiator proteins (DnaA, Gene II and Gene A) was the focus of this research. All these proteins have presented problems associated with obtaining them in large quantities for structural and biochemical studies.

High-level overproduction of DnaA was achieved by changing the GUG start codon to AUG, replacing the natural RBS with a synthetic RBS perfectly complementary to the 3′ end of 16S rRNA and placing the gene under the control of strong λ promoters. The levels of overproduced DnaA represented 84 000 copies per cell. A new strategy for large-scale purification was developed following identification of conditions ($MgCl_2$ and ATP) in which the protein was soluble at moderate ionic strength. Highly pure (> 95%), fully active, monomeric DnaA was obtained in an overall yield of 25%. It was shown that less than two molecules of DnaA (purified in this manner) was required to replicate one circle of M13 A-site DNA.

The DnaA protein was shown to be monomeric in solution based on sedimentation equilibrium studies. The monomeric molecular mass was accurately determined by electrospray mass spectrometry (52 422.5 Da) which was within 2.8 Da of the theoretical mass calculated from the amino acid sequence. A second component was observed in the spectrum that was 79.9 Da larger. This suggested that the protein was either phosphorylated or sulfated. *In vitro* phosphorylation of DnaA with ^{32}P using crude cell extracts from a *dnaA* mutant strain produced labelled protein on an autoradiograph that corresponded to DnaA, providing evidence of putative phosphorylation. Analysis of tryptic fragments by electrospray mass spectrometry have shown that the site of putative phosphorylation resides within the N-terminal 107 residues. A disulfide linkage has been identified in the region (between residues 8 and 66), protecting it from tryptic digestion.

M13 *gene II* was cloned into λ and T7 promoter vectors and high-level overproduction was achieved (20-30% of the total cell protein). The overproduced Gene II was insoluble and required 6 M guanidine.HCl or 8 M urea for solubilisation. The protein remained soluble in high pH solutions when the urea concentration was reduced A procedure for Gene II purification was devised and 30-60 mg of highly pure (> 95%) protein was obtained. Gene II protein was active in relaxation of M13 RFI DNA but was inactive in RF replication. It is possible that an unknown factor, removed during the purification, is required for RF replication.

Proteins tagged with a poly-histidine tail can be purified in a one-step procedure using metal chelate [Ni(II)-NTA] chromatography. Expression vectors were constructed to tag protein at the N-terminus with six histidines residues and the level of overproduction of his_6-Gene II was similar to that obtained from the wild type protein. His_6-Gene II was bound to Ni(II)-NTA resin under denaturing conditions then the urea was diluted from 8 to 1 M over several hours. Pure, soluble protein was eluted from the column. However, the protein was inactive in relaxation of M13 RFI DNA. This suggests that the protein was probably incorrectly folded.

Several fragments of ϕX174 *gene A* and *gene A** were cloned but attempts to clone the intact gene were unsuccessful. Investigation of the nucleic acid sequence of *gene A* revealed a putative promoter, located 57 bp upstream from the start of *gene A**. The -10, -35 regions and mRNA start sites of this promoter have high homology to the consensus sequences of these elements, spaced in accordance with a functional promoter. During the cloning of *gene A*, *gene A** may be expressed constitutively from its own promoter compromising the experiments by either having a lethal effect on cell growth by shutoff of cell replication or cleavage of the ϕX174 (+) strand origin, effecting plasmid replication.

CONTENTS

CHAPTER 1

INTRODUCTION

The genetic material of all organisms must be faithfully replicated once (and only once) per cell cycle. Some of the replication machinery is also used for DNA recombination and repair processes. *Escherichia coli*, which contains a single, circular genome of ~4 700 kb has been used as a model for Gram-negative bacteria. The current understanding of replication has been based on the identification, genetic and biochemical analysis, and function of > 30 protein components involved in the replicative processes of bacteriophages which rely on host proteins for replication and several plasmids which rely solely on the chromosomal origin for replication. In *E. coli*, replication starts at the origin of replication (*oriC*) and synthesis occurs bidirectionally, terminating at a site ~180° from the origin. On the basis of this information DNA replication can be divided into three separable stages: initiation, elongation and termination.

Chromosomal replication is controlled at the level of initiation, primarily by the action of the DnaA replication initiator protein at the origin of replication, and its role in the regulation of its own and other genes. The *E. coli* bacteriophages M13 and ϕX174 rely on their host for the supply of all replication machinery with the exception of the phage-encoded replicator initiator proteins. These (Gene II of M13 and Gene A of ϕX174) are endonucleases which initiate DNA replication in a manner completely different to that of DnaA in initiation of chromosomal replication. The replication initiator proteins DnaA (*E. coli*), Gene II (M13) and Gene A (ϕX174) are the focus of work presented in this thesis.

1.1 Replication of the *Escherichia coli* Chromosome

1.1.1 The E. coli Origin of Replication, oriC

Initiation of DNA replication of the *E. coli* chromosome occurs at the origin of replication (*oriC*) located at 83.5 min on the genetic map (Bird *et al.*, 1972; Louarn *et*

al., 1974; von Meyenburg *et al.*, 1977). The minimal origin region was identified by determining the smallest *oriC* segment that could sustain replication of a ColE1 plasmid in a DNA polymerase I temperature-sensitive host. DNA polymerase I is required for leading strand synthesis of the Col E1 replicon, but not for the host. Thus, replication of the plasmid at a non-permissive temperature indicated an active *oriC* (Oka *et al.*, 1980). A minimal *oriC* DNA segment of 245 bp was capable of autonomous DNA replication (Oka *et al.*, 1980).

Genetic manipulation and biochemical studies have revealed several interesting features within *oriC* (Figure 1.1) (Oka *et al.*, 1982; Asada *et al.*, 1982; Oka *et al.*, 1984). Four 9-mer repeat sequences [5´-TTAT($^{C}/_{A}$)CA($^{C}/_{A}$)A], R1-R4, and a fifth sequence [5´-TCATTACA, designated M] are binding sites for DnaA (Fuller *et al.*, 1984; Woelker and Messer, 1993), the protein that initiates DNA replication (Figure 1.1). There are three 13-mer AT-rich repeat sequences (denoted L, M and R) in an AT-rich cluster located at the left edge of *oriC* and these play a role in the localised melting of the duplex DNA during initiation to allow the loading of the DnaB helicase. The sequence 5´-GATC occurs 11 times in *oriC* and is the recognition site for adenine methylation by deoxyadenosine (Dam) methyltransferase. Methylation of these sites is thought to play an important role in replication *in vivo* and *in vitro* (Hughes *et al.*, 1984; Messer *et al.*, 1985; Smith *et al.*, 1985). Binding sites for DNA gyrase (Lother *et al.*, 1984), the DnaBC complex (Kornberg *et al.*, 1987), IHF (Polaczek, 1990; Filutowitz and Roll, 1990), Fis (Gille *et al.*, 1991; Filutowitz *et al.*, 1992) and SeqA (Lu *et al.*, 1994; Slater *et al.*, 1995) are located in *oriC*.

The *oriC* sequence is very highly conserved in diverse enteric bacteria and its origins on plasmids in *E. coli* can be replicated autonomously (Figure 1.2) (Zyskind *et al.*, 1983). Therefore, the sequence elements must all be recognised by *E. coli* replication proteins. In fact the DnaA proteins of *S. typhimurium* and *S. marcescens* have been shown to function in *E. coli* (Skovgaard and Hansen, 1987). The origins of more

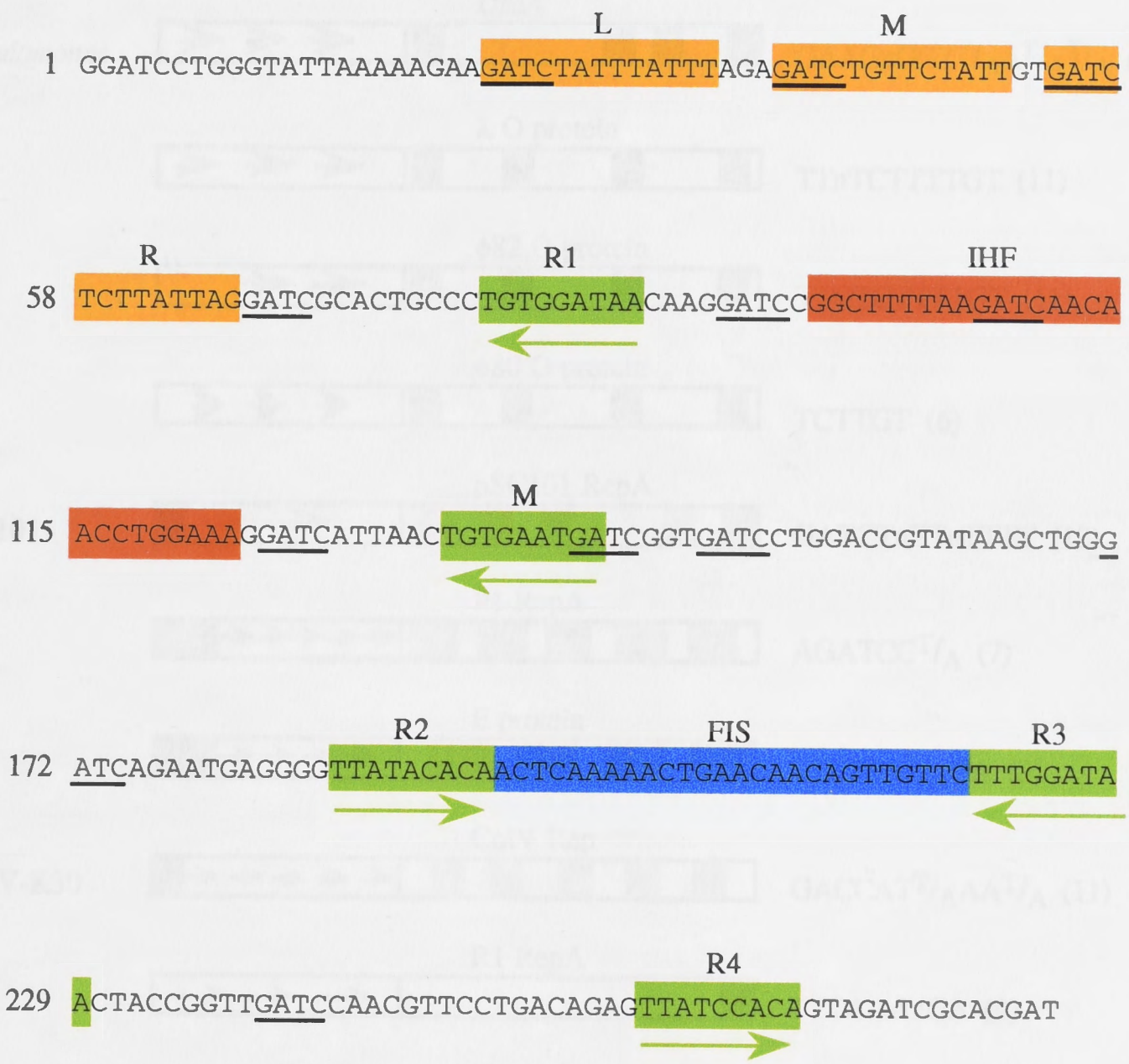


Figure 1.1

The nucleotide sequence of the minimal *oriC* region of *E. coli* chromosome. The DnaA boxes (green), 13-mer AT-rich repeats (yellow), IHF binding site (red), FIS binding site (blue) and *dam* methylation sites (underlined in black) are shown. The orientations of the DnaA boxes is indicated by the arrows. This diagram was adapted from Woelker and Messer (1993).

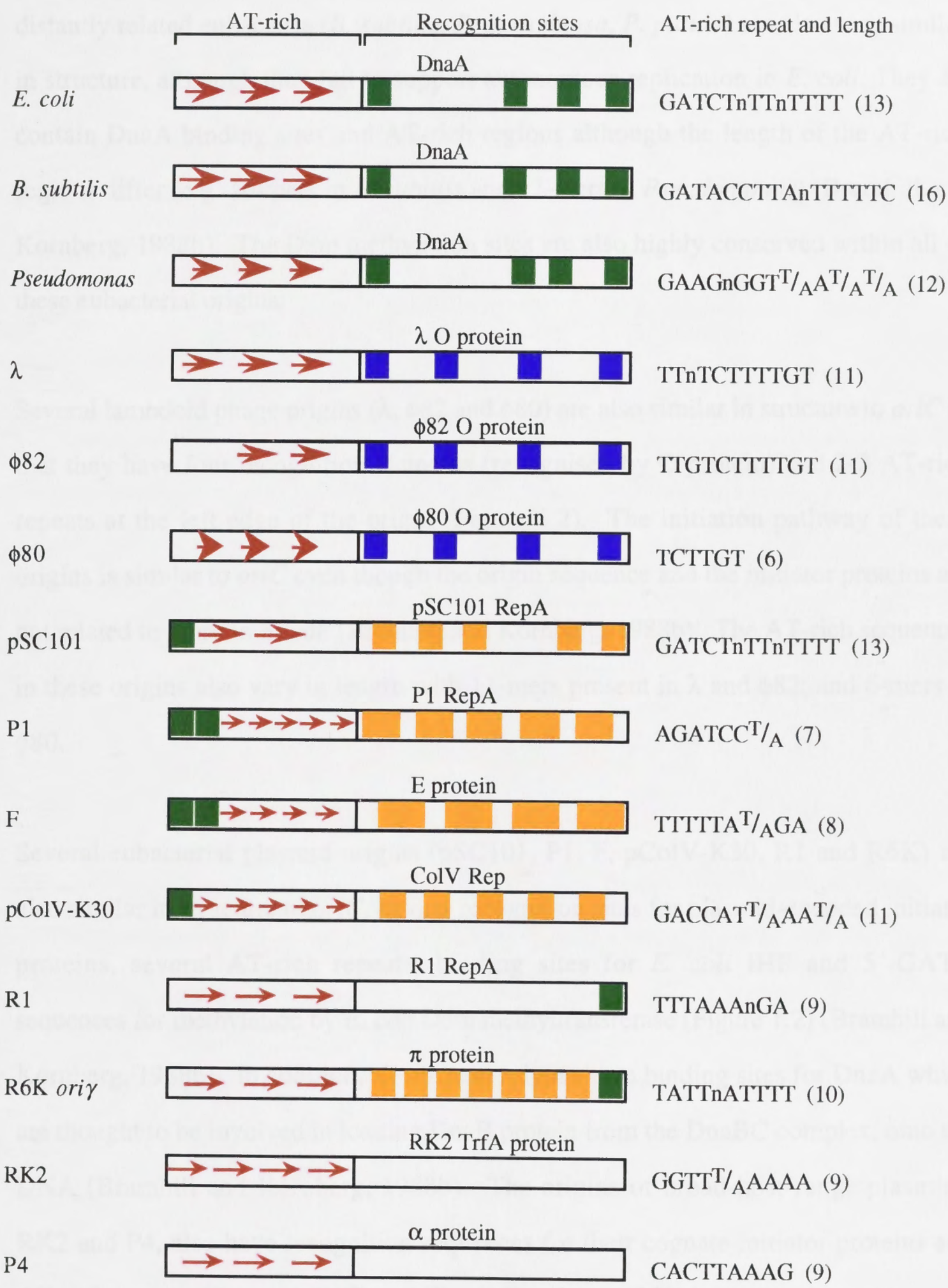


Figure 1.2

A comparison of the origin regions of several bacteria, lambdoid phages and prokaryotic plasmids showing the number, sequence and length of AT-rich repeats, the recognition proteins and the number of sites to which they bind [adapted from Bramhill and Kornberg (1988b), and Kornberg and Baker (1992)]. DnaA boxes, λO protein binding sites and plasmid recognition protein binding sites are represented by green, blue and yellow boxes, respectively. The red arrows indicate AT-rich repeats.

distantly related eubacteria (*B. subtilis*, *P. aeruginosa*, *P. putida*) are also very similar in structure, although they fail to support autonomous replication in *E. coli*. They all contain DnaA binding sites and AT-rich regions although the length of the AT-rich regions differ (e.g. 16-mers in *B. subtilis* and 12-mers in *Pseudomonas*) (Bramhill and Kornberg, 1988b). The Dam methylation sites are also highly conserved within all of these eubacterial origins.

Several lambdoid phage origins (λ, ϕ82 and ϕ80) are also similar in structure to *oriC* in that they have four recognition elements (recognised by O protein) and 2-3 AT-rich repeats at the left edge of the origin (Figure 1.2). The initiation pathway of these origins is similar to *oriC* even though the origin sequence and the initiator proteins are not related to those in *E. coli* (Bramhill and Kornberg, 1988b). The AT-rich sequences in these origins also vary in length with 11-mers present in λ and ϕ82, and 6-mers in ϕ80.

Several eubacterial plasmid origins (pSC101, P1, F, pColV-K30, R1 and R6K) are also similar in structure to *oriC*, having recognition sites for plasmid-encoded initiator proteins, several AT-rich repeats, binding sites for *E. coli* IHF and 5′-GATC sequences for methylation by *E. coli* Dam methyltransferase (Figure 1.2) (Bramhill and Kornberg, 1988b). In addition, all of these origins have binding sites for DnaA which are thought to be involved in loading DnaB protein from the DnaBC complex, onto the DNA (Bramhill and Kornberg, 1988b). The origins of broad-host-range plasmids, RK2 and P4, also have recognition sequences for their cognate initiator proteins and AT-rich regions (Figure 1.2) (Bramhill and Kornberg, 1988b).

The similarities observed among all these origins suggest that the mechanism of initiation of DNA replication in eubacteria, lambdoid phages and several plasmids is also very similar.

1.1.2 Initiation

The mechanism for initiation of DNA replication can be separated into several stages: recognition of the origin, initial melting of the duplex DNA, and assembly of the prepriming complex (see Figure 1.3).

Recognition of the origin

This stage involves the cooperative binding of 10-30 DnaA monomers to the four 9-mer DnaA boxes within *oriC* which lead to the formation of a nucleoprotein complex (Fuller *et al.*, 1984; Woelker and Messer, 1993). DnaA in either the ADP- or ATP-bound form can bind to *oriC*, although only the ATP-bound form is active in initiation. The nucleoprotein structure formed appears spherical with the DNA wrapped around the periphery of the protein, as revealed by electron microscopy (Fuller *et al.*, 1984; Funnel *et al.*, 1987). Approximately 200 bp of the origin are protected from DNase I digestion. This did not include the AT-rich 13-mers (Fuller *et al.*, 1984). Complex formation occurs at low temperature (< 30°C).

Initial melting of the duplex DNA

A temperature-dependent structural change in the DnaA-*oriC* complex promotes localised melting of the DNA in the AT-rich 13-mers positioned at the leftmost edge of the origin (Bramhill and Kornberg, 1988a). The open complexes were identified by this region being susceptible to cleavage by the single-strand-specific endonuclease P1 (Bramhill and Kornberg, 1988a). The opening occurs sequentially from the right 13-mer leftward and although DNase I protection was not observed, DnaA interaction with the 13-mers has been implicated (Yung and Kornberg, 1989). The template must be negatively supercoiled (Baker and Kornberg, 1988) and small amounts of HU or IHF proteins stimulate the reaction, presumably by stabilising a desirable bend in the DNA (Skarstad *et al.*, 1990).

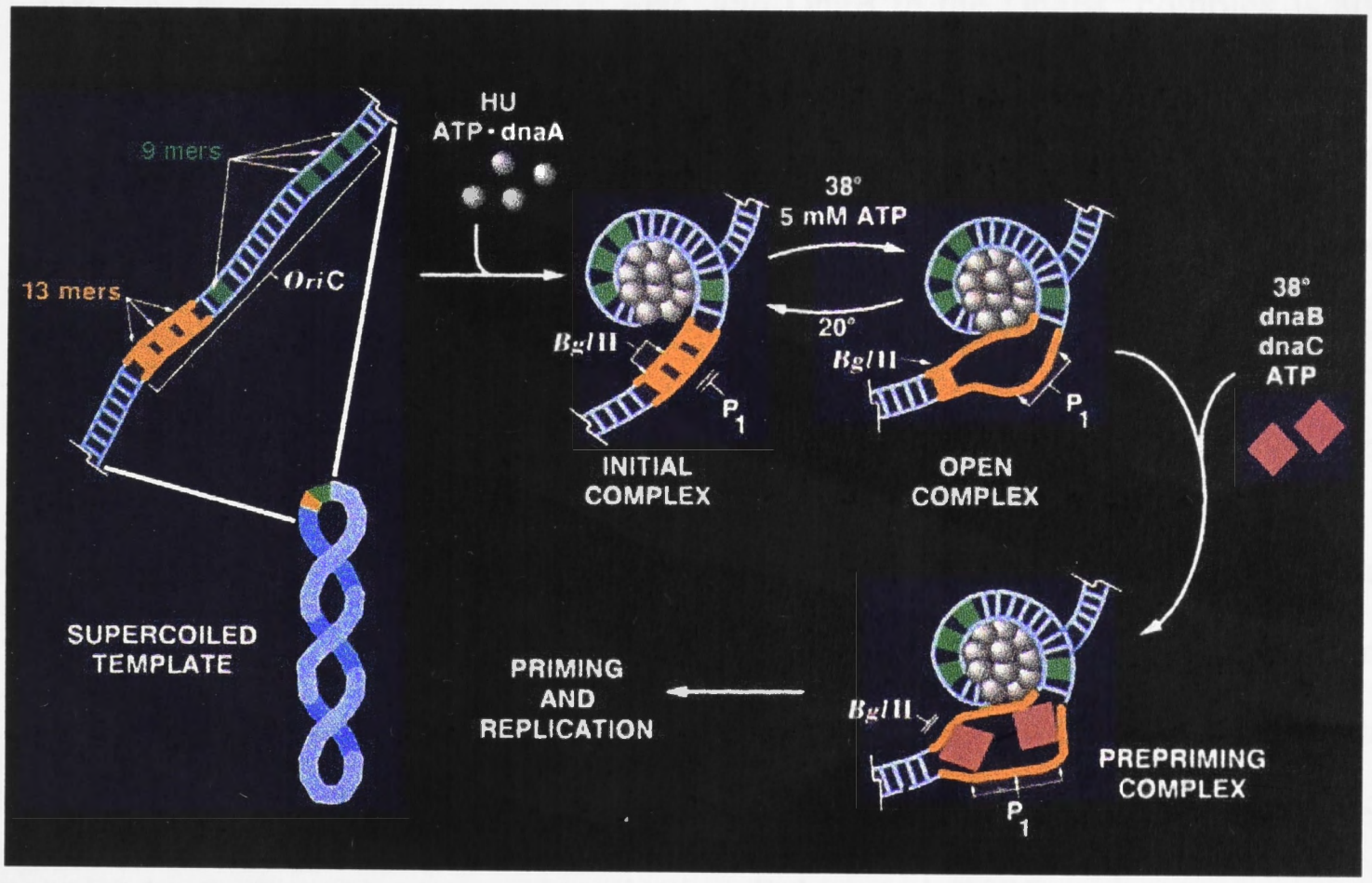


Figure 1.3

Model for initiation of *E. coli* chromosomal DNA replication *in vitro*. The schematic diagram does not show the stoichiometry of the proteins in the postulated complexes (from Kornberg and Baker, 1992). Details are given in Section 1.1.2.

Prepriming complex formation

The DnaB helicase in the presence of ATP forms a complex with the DnaC protein (Wichner and Hurwitz, 1975; Kobori and Kornberg, 1982; Wahle *et al.*, 1989a, 1989b). It has been proposed that DnaC in the DnaBC complex recognises DnaA, facilitating the binding of DnaB to the single-stranded DNA in the open complex at the origin and stimulating the formation of a prepriming complex (Funnel *et al.*, 1987; Wahle *et al.*, 1989a; Masai *et al.*, 1990b; Hwang and Kornberg, 1992). Direct interaction between DnaA bound to *oriC* and DnaB in the DnaBC complex is also required to permit binding of DnaB (Marszalek and Kaguni, 1994). DnaC is required for the assembly but is not stably maintained in the prepriming complex as it inhibits the helicase activity of DnaB and must be removed for replication to proceed (Allen and Kornberg, 1991; Kornberg and Baker, 1992). The prepriming complex containing DnaA, DnaB and (probably) HU protein bound to *oriC* is stable and isolable. It is larger than the initiation complex and encompasses an additional 50 bp in the region of the 13-mers (Funnel *et al.*, 1987). Further melting in the 13-mer region of the complex was detected by an increased sensitivity to endonuclease P1 (Bramhill and Kornberg, 1988a). Once positioned on the single-stranded DNA, the DnaB helicase migrates bidirectionally from the DnaA-*oriC* complex to provide a template for subsequent priming by the DnaG primase and replication (Baker *et al.*, 1986).

The initiation of bacteriophage λ DNA replication has been reported to function in an analogous manner to initiation at *oriC* (Figure 1.4) (LeBowitz *et al.*, 1985; Alfano and McMacken, 1989a; Dodson *et al.*, 1989; Liberek *et al.*, 1990). Replication is initiated at *oriλ* which is structurally similar to *oriC* (see Section 1.1). The λO and P proteins are functional analogs of *E. coli* DnaA and DnaC, respectively. λO protein binds to four 18-mer direct repeats (each of which is an inverted repeat) within *oriλ*, melting the duplex DNA in the vicinity of the adjacent AT-rich repeats (Furth and Wickner, 1983). The λO protein was shown to bind to the 18-mer repeat (Tsurimoto and Matsubara, 1981; Zahn and Blattner, 1985), most likely as a dimer (Wickner and Zahn, 1986).

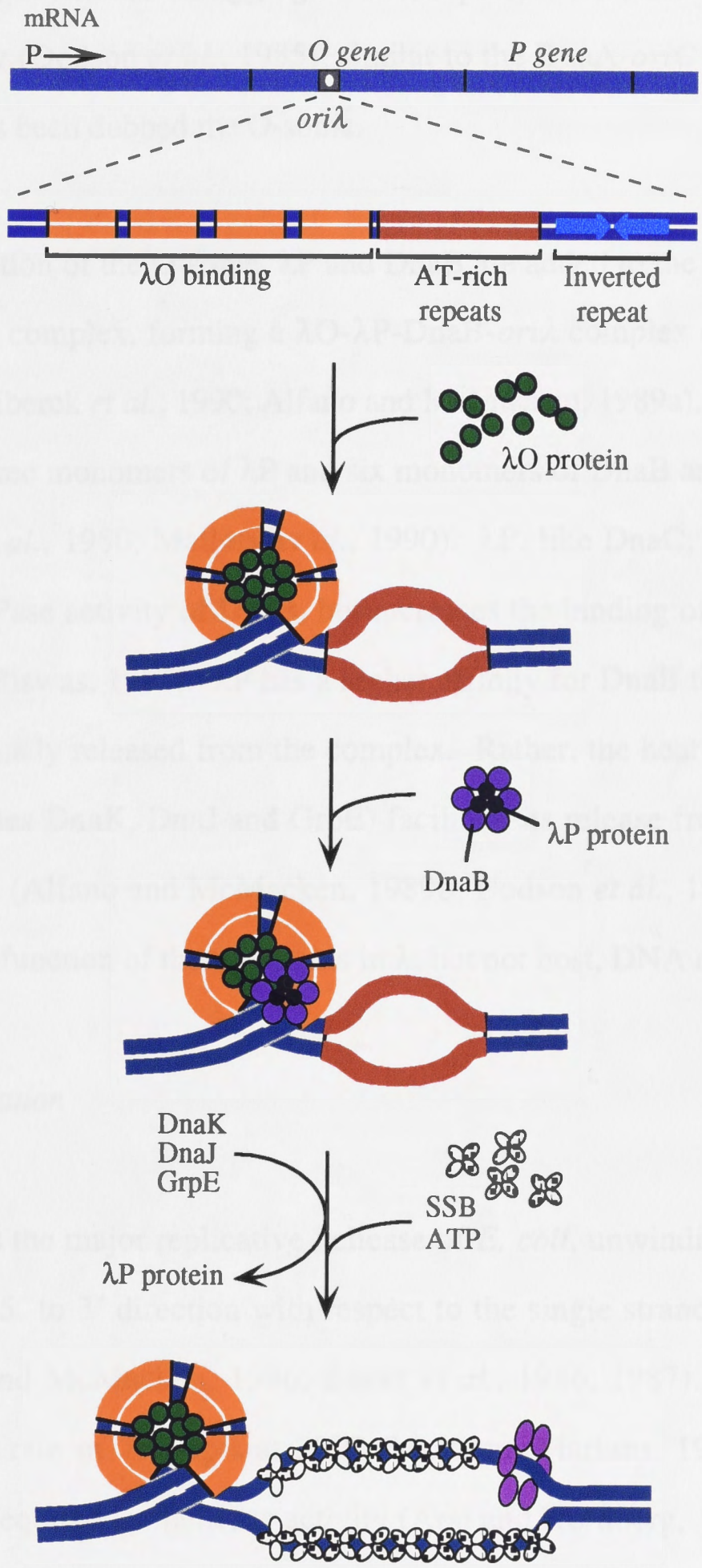


Figure 1.4

A model for the initiaton of DNA replication at *oriλ*. This process is described in detail in Section 1.1.2 (Figure courtesy of Dr N.E. Dixon).

Electron microscopy identified an aggregate of λO protein with the DNA wrapped around the periphery (Dodson *et al.*, 1985), similar to the DnaA-*oriC* complex. The λO-*oriλ* complex has been dubbed the O-some.

Following the formation of the O-some, λP and DnaB are added to the complex, most likely as a λP-DnaB complex, forming a λO-λP-DnaB-*oriλ* complex (Dodson *et al.*, 1985; 1986; 1989; Liberek *et al.*, 1990; Alfano and McMacken, 1989a). The λP-DnaB complex contains three monomers of λP and six monomers of DnaB and the complex is isolable (Klein *et al.*, 1980; Mallory *et al.*, 1990). λP, like DnaC, suppresses the ATP-dependent ATPase activity of DnaB, but increases the binding of ATP to DnaB 6-fold (Biswas and Biswas, 1987). λP has a higher affinity for DnaB than DnaC does and is not spontaneously released from the complex. Rather, the heat shock proteins (molecular chaperones DnaK, DnaJ and GrpE) facilitate its release from the λO-λP-DnaB-*oriλ* complex (Alfano and McMacken, 1989b; Dodson *et al.*, 1986,1989), and this accounts for the function of these proteins in λ, but not host, DNA replication.

1.1.3 Elongation

The DnaB protein is the major replicative helicase of *E. coli*, unwinding the parental duplex DNA in the 5′ to 3′ direction with respect to the single strand on which it is bound (LeBowitz and McMacken, 1986; Baker *et al.*, 1986, 1987). DnaB duplex melting occurs at a rate of 730 bp/s at 30°C (Mok and Marians, 1987, Wu *et al.*, 1992a) and ATP is required for helicase activity (Arai and Kornberg, 1981a). During the elongation stage, the multifunctional DnaB helicase is also involved in assembly, either directly or indirectly, of the remainder of the replication proteins at the replication fork (Baker and Wickner, 1992).

The DnaB helicase, when positioned on the single-stranded region of *oriC*, proceeds to unwind the duplex DNA in the vicinity of the origin (Baker *et al.*, 1986). SSB protein

binds cooperatively to the single-stranded DNA generated by the helicase activity of DnaB, preventing the reassociation of the complementary strands (Chase and Williams, 1986). DnaB then activates the synthesis of short RNA primers in the region of *oriC*, probably by forming a seconday structure in the DNA template that can be utilised by the DnaG primase (Arai and Kornberg, 1981b). This initial priming probably occurs in the single-stranded region of *oriC* in the presence of primase and replication proteins DnaA, DnaB, DnaC and DNA gyrase (van der Ende *et al.*, 1985). It has been demonstrated that a direct interaction between the DnaB helicase and DnaG primase is essential for primer synthesis at the replication fork (Marians, 1992; Wu *et al.*, 1992b; Tougu *et al.*, 1994). RNA primers are necessary for the replication of the parental template by DNA polymerase III holoezyme, the major replicative DNA polymerase of *E. coli* (McHenry, 1988). The DNA polymerase III holoenzyme is unable to synthesise DNA *de novo* but can extend chains that have a 3′-^{OH} terminus of a nucleotide paired with the template strand. Replication occurs in the 5′ to 3′ direction on both strands of the duplex in a semidiscontinuous manner, whereby one strand is synthesised continuously (leading strand) and the other discontinuously (lagging strand) in short segments termed Okazaki fragments (Wu *et al.*, 1992a; Zechner *et al.*, 1992a, 1992b). The Okazaki fragments are 1000-2000 bp in length and are joined by DNA ligase after removal of the initiator RNA to form a continuous strand of DNA (Wu *et al.*, 1992a; Zechner *et al.*, 1992a, 1992b). The nature of the discontinuous strand necessitates the involvement of several other proteins for the removal of the RNA primers, filling of gaps and ligation for conversion to a continuous strand. DNA polymerase I and DNA ligase perform these functions, and RNaseH may also be involved (Ogawa and Okazaki, 1984; Funnel *et al.*, 1986).

DnaG primase can only act alone under certain circumstances to synthesise RNA primers (e.g. on phage G4 ssDNA template) (Bouché *et al.*, 1978; Masai *et al.*, 1990a). In most cases it requires DnaB which functions as a "mobile promoter" of primer synthesis (Baker and Wickner, 1992). For general priming that occurs on

ssDNA templates in the absence of SSB, only DnaB and DnaG primase (in the presence of ATP) are required for synthesis of RNA primers (Figure 1.5) (Arai and Kornberg, 1979). The association of DnaB and DnaG primase with DNA is distributive during general priming. The proteins continually dissociate and reassociate with the template (Lehman and Kornberg, 1992). When the binding of DnaB limits the rate of priming, DnaC stimulates the binding of DnaB to the DNA 5-fold (Wahle *et al.*, 1989b). ATP hydrolysis that drives the helicase activity of DnaB is not required for general priming (Baker and Kornberg, 1992). However, when the ssDNA template is completely coated with SSB, general priming is inhibited (Arai and Kornberg, 1979; LeBowitz and McMacken, 1986).

Both DnaB and DnaG primase are required for the replication of the *E. coli* chromosome and the interaction between the proteins is essential, as primase alone is inert in the presence of ssDNA (Lehman and Kornberg, 1992). The direct interaction of DnaB and DnaG primase, demonstrated recently, is necessary for introducing primase to the replication fork (Tougu *et al.*, 1994). However, the mechanism by which DnaB activates DnaG primase is unclear. One notion is that it produces secondary structural elements that are recognised by primase (Arai and Kornberg, 1981a).

As DnaB and DnaG primase are unable to assemble on either SSB-coated ssDNA or associate with duplex DNA, specific mechanisms for primosome assembly and priming must be required. Studies on the conversion of ssDNA of several single-stranded bacteriophages (M13, G4 and ϕX174) to replicative form (RF) (ss→RF) have revealed several priming mechanisms. Priming of M13 is performed by the host RNA polymerase (Figure 1.5) (Brutlag *et al.*, 1971). In the priming of G4 ssDNA coated by SSB, primase alone was able to synthesise an RNA primer via the formation of a specific SSB-dependent secondary structure on the single-stranded template (Figure 1.5) (Lambert *et al.*, 1986; Hiasa *et al.*, 1990).

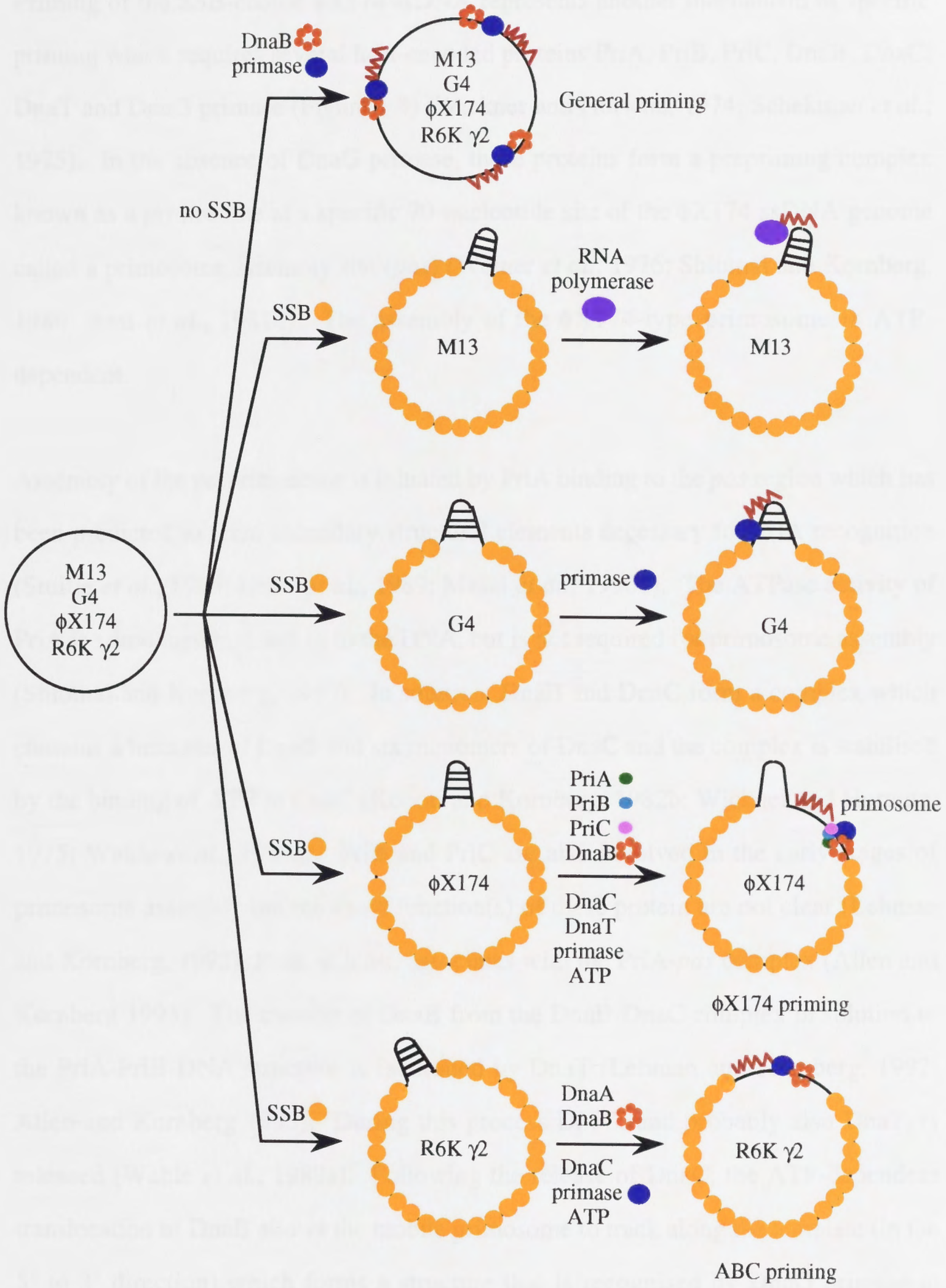


Figure 1.5

The mechanisms for general (nonspecific) priming by DnaB and primase of ssDNA in the absence of SSB, and specific priming of SSB-coated M13, G4, ϕX174 and R6K γ2 DNAs. These mechanisms are described in detail in Section 1.1.3. This figure was adapted from Kornberg and Baker (1992).

Priming of the SSB-coated ϕX174 ssDNA represents another mechanism of specific priming which requires several host-encoded proteins PriA, PriB, PriC, DnaB, DnaC, DnaT and DnaG primase (Figure 1.5) (Wickner and Hurwitz, 1974; Schekman *et al.*, 1975). In the absence of DnaG primase, these proteins form a prepriming complex known as a primosome at a specific 70-nucleotide site of the ϕX174 ssDNA genome called a primosome assembly site (*pas*) (Weiner *et al.*, 1976; Shlomai and Kornberg, 1980; Arai *et al.*, 1981d). The assembly of the ϕX174-type primosome is ATP-dependent.

Assembly of the preprimosome is initiated by PriA binding to the *pas* region which has been predicted to form secondary structural elements necessary for PriA recognition (Stuitje *et al.*, 1984; Hiasa *et al.*, 1989; Masai *et al.*, 1990b). The ATPase activity of PriA is stimulated by binding to the DNA, but is not required for primosome assembly (Shlomai and Kornberg, 1980). In solution, DnaB and DnaC form a complex which contains a hexamer of DnaB and six monomers of DnaC and the complex is stabilised by the binding of ATP to DnaC (Kobori and Kornberg, 1982b; Wickner and Hurwitz, 1975; Wahle *et al.*, 1989a). PriB and PriC are also involved in the early stages of primosome assembly but the exact function(s) of these protein are not clear (Lehman and Kornberg, 1992), PriB, at least, associates with the PriA-*pas* complex (Allen and Kornberg 1993). The transfer of DnaB from the DnaB-DnaC complex in solution to the PriA-PriB-DNA structure is facilitated by DnaT (Lehman and Kornberg, 1992; Allen and Kornberg 1993). During this process DnaC, and probably also DnaT, is released (Wahle *et al.*, 1989a). Following the release of DnaC, the ATP-dependent translocation of DnaB allows the mobile primosome to track along the template (in the 5′ to 3′ direction) which forms a structure that is recognised by DnaG primase at numerous positions (Wahle *et al.*, 1989a). Primase is not required for either stability or mobility of the primosome and may continually associate and dissociate to synthesise RNA primers (Wu *et al.*, 1992b).

The final composition of the primosome is still to be determined (Lehman and Kornberg, 1992). However, it has been reported (Lee and Marians, 1989) that the ϕX174-type primosome could act as a helicase in both directions. Thus, DnaB (5′ to 3′ helicase) and PriA (3′ to 5′ helicase) would be expected to be stable components in the primosome complex. Allen and colleagues (Allen *et al.*, 1993; Allen and Kornberg, 1993) have shown experimentally that PriB remains associated in the primosome.

A further mechanism of priming has been observed by Masai and colleagues (1990a, b). This involves a different mode of loading DnaB onto a ssDNA template *in vitro* and may mimic one aspect of initiation at *oriC*. The ssDNA template (coated with SSB) has a specific hairpin structure carrying a single DnaA box derived from the plasmid R6K γ-origin region. Priming requires only DnaA, DnaB, DnaC and primase and has been denoted ABC-priming (Figure 1.5) (Masai *et al.*, 1990a).

DnaA initiates ABC-priming by recognition and binding to its cognate site on the (double-stranded) stem of the hairpin structure, and this permits the loading of DnaB from the DnaBC complex. A prepriming complex is formed which contains DnaA and DnaB bound to the DNA template. The helicase activity of DnaB is inhibited in the prepriming complex. Upon addition of DnaG primase the complex is converted to a complete ABC-primosome (Masai *et al.*, 1990a). The ABC primosome in conjunction with DNA polymerase III holoenzyme can efficiently convert ssDNA into a double-stranded replicative form (Masai *et al.*, 1990a). The assembly of the ABC-primosome is thought to model the DnaA-dependent assembly of replication proteins at *oriC* (Masai *et al.*, 1990a; Marians, 1992).

Both the ABC- and ϕX174-type mechanisms for loading DnaB have been shown to be capable of leading to priming and subsequent leading- and lagging-strand DNA synthesis (Arai *et al.*, 1981b; Arai and Kornberg, 1981d; Masai and Arai, 1989; Masai *et al.*, 1990a; Wu *et al.*, 1992a). ABC-priming shows that PriA, PriB, PriC and DnaT

are not required to attract DnaG primase to the replication fork (Wu *et al.*, 1992a). However, DNA replication is unstable in strains carrying *dnaT*-^{ts} mutation (Masai et al., 1986; Masai and Arai, 1988) and cell viability is reduced in strains which carry a *priA* deletion (Lee and Kornberg, 1991; Nurse *et al.*, 1991). It has been proposed (Nurse *et al.*, 1991) that although a ϕX174-type primosome is not involved in assembly of a primosome at *oriC*, completion of chromosomal replication could be dependent on subsequent assembly of ϕX174-type primosomes. This may occur if nucleoprotein structures migrating from *oriC* stall or dissociate (Marians, 1992). A role for ϕX174-type primosomes in *E. coli* replication has been demonstrated by Allen and colleagues (Allen *et al.*, 1993; Allen and Kornberg, 1993). The presence of *pas* sequences on the chromosome and the ability of the ϕX174-type primosome to transfer DnaB from one strand to the other (Allen *et al.*, 1993), suggested its possible role in reinitiation of DNA synthesis when replication forks encounter difficulties distant from *oriC*. Under the appropriate conditions, the *pas* sequence and the ϕX174-type primosome could initiate bidirectional replication of duplex DNA (Allen *et al.*, 1993; Allen and Kornberg, 1993). These facts indicate that both ABC and ϕX174 primosome-mediated priming may play important roles in *E. coli* DNA replication.

The major replicative DNA polymerase in *E. coli* is the DNA polymerase III holoenzyme, an asymmetric complex of ten subunits (α, ε, θ, τ, γ, δ, δ′, χ, ψ, β) (McHenry, 1988; Maki *et al.*, 1988). The holoenzyme is highly processive, synthesising DNA at a rate of ~500 nucleotides per second at 30°C during replication of a singly-primed ϕX174 ssDNA template *in vitro* (O'Donnell and Kornberg, 1985). The multisubunit complex is responsible not only for high-fidelity replication of the *E. coli* chromosome but also has the ability to proof-read the newly synthesised daughter strand (McHenry, 1988; Maki *et al.*, 1988). The genes encoding all ten subunits have now been cloned and overexpressed, and this has resulted in exciting progress in determination of how the subunits assemble and their functional roles in the multisubunit holoenzyme (see Kelman and O'Donnell, 1995).

Several proteins (PriA, PriB, DNA gyrase), together with DnaB helicase, DnaG primase, DNA polymerase III holoenzyme and some other additional components, presumably form a large nucleoprotein structure at a replication fork called a replisome (Kornberg and Baker, 1991; Lehman and Kornberg, 1992). The replisome operation at the replication fork is expected to effectively coordinate replication of both the continuously- and discontinuously-synthesised strands (Figure 1.6).

1.1.4 Termination

Replication forks are blocked within a terminus region which spans ~400 kb roughly 180° from *oriC* (de Massy *et al.*, 1987; Hill *et al.*, 1987). The arrest of replication fork progression is attributed to the specific binding of Tus (terminator utilisation substance) protein to a nonpalindromic *Ter* sequence (Hidaka *et al.*, 1988; Pelletier *et al.*, 1989; Hill *et al.*, 1989). The *Ter* sequences are 22-23 bp and the Tus-*Ter* complex acts in an orientation-dependent manner (Hill *et al.*, 1988; Hidaka *et al.*, 1988). Six *Ter* sites in *E. coli* are clustered at the left and right hand edges of the terminus region (Hill *et al.*, 1988; Hidaka *et al.*, 1988, 1991; Sharma and Hill, 1992; Kuempel *et al.*, 1989). Three (*TerA*, *TerD* and *TerE*) are oriented to arrest a counterclockwise-moving replication fork and three (*TerB*, *TerC* and *TerF*) to arrest a clockwise-moving replication fork. This terminus region acts as a replication fork trap, whereby replication forks are permitted to enter but are prohibited from leaving (Kuempel *et al.*, 1989; Baker, 1995). Experimental evidence suggests the DnaB helicase is inhibited by Tus bound to the appropriately orientated *Ter* site (Khatri *et al.*, 1989; Lee *et al.*, 1989), and correctly orientated Tus-*Ter* complexes have also been shown to stall ϕX174 primosomes (Baker *et al.*, 1995).

However, termination of replication by the Tus-mediated mechanism is not required for *E. coli*. When the terminus region is deleted, cell growth and viability under experimental conditions are not affected (Kobayashi *et al.*, 1989; Hill *et al.*, 1989;

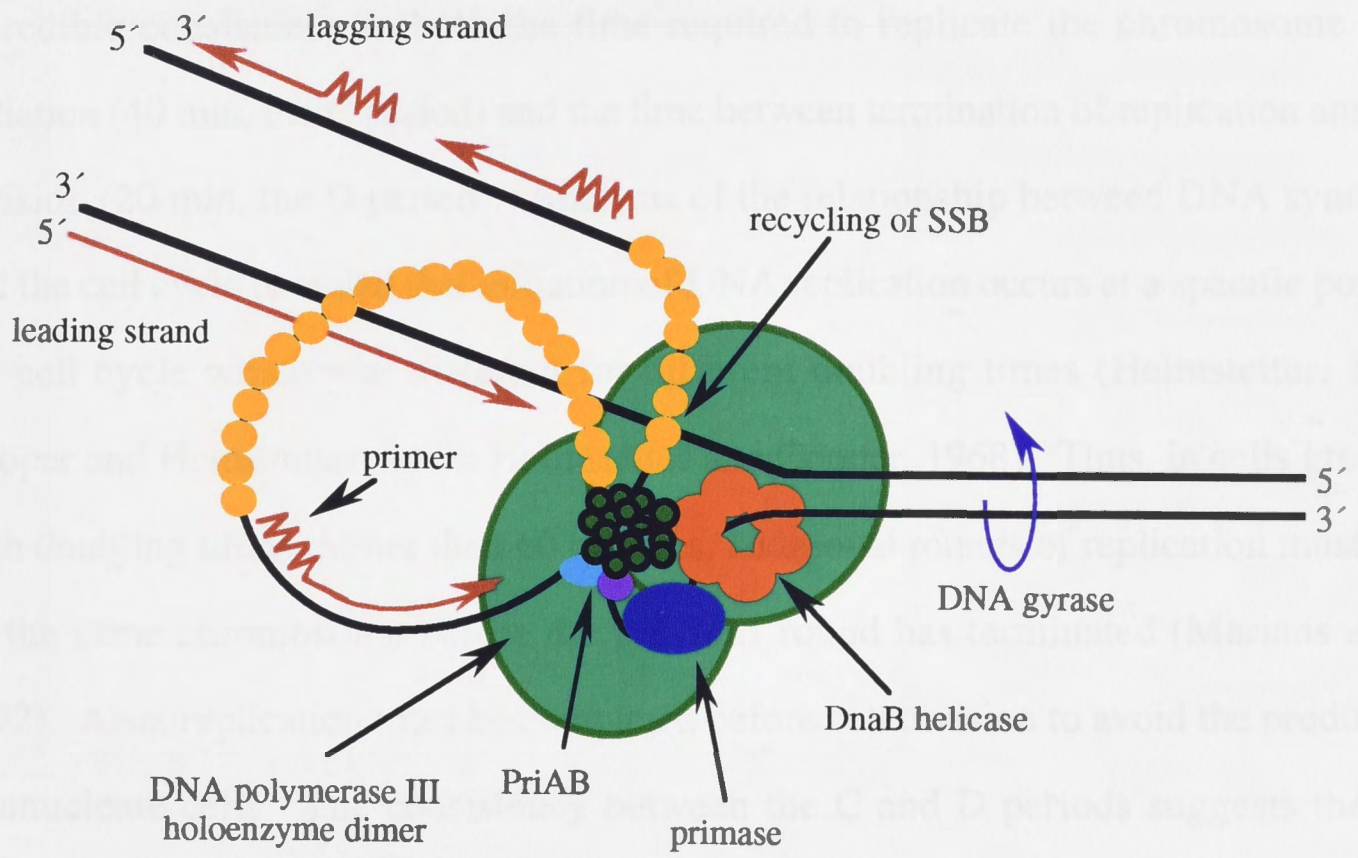


Figure 1.6

A hypothetical scheme for concurrent lagging and leading strand replication via a nucleoprotein complex called a replisome (Kornberg and Baker, 1992). The exact composition of the replisome is not known although it probably contains DNA polymerase III holoenzyme, the primosome (containing DnaB and PriA helicases, PriB protein and DnaG primase), SSB, DNA gyrase, DNA ligase and DNA polymerase I (details are in Section 1.1.3) (Figure courtesy of Dr N.E. Dixon).

Roecklein *et al.*, 1991). In this situation, replication is apparently halted by the collision of the opposing replication forks (Lehman and Kornberg, 1992; Baker and Wickner, 1992).

1.1.5 Regulation, Timing and Synchrony of DNA Replication Initiation

E. coli cells, growing with doubling times between 20 and 60 minutes, exhibit incredible consistency in both the time required to replicate the chromosome after initiation (40 min, the C period) and the time between termination of replication and cell division (20 min, the D period). Analysis of the relationship between DNA synthesis and the cell cycle revealed that initiation of DNA replication occurs at a specific point in the cell cycle which was different for different doubling times (Helmstetter, 1967; Cooper and Helmstetter, 1968; Helmstetter and Cooper, 1968). Thus, in cells growing with doubling times shorter than 60 minutes, additional rounds of replication must start on the same chromosome before the previous round has terminated (Marians *et al.*, 1992). Also, replication must be completed before cell division to avoid the production of anucleate cells. The consistency between the C and D periods suggests that the timing of initiation must be precisely regulated during the cell cycle (Marians *et al.*, 1992). It has been proposed (Donachie, 1968) that new rounds of replication are initiated at a ratio of chromosomal origins to cell mass which is rather constant and independent of cell growth rate. This could be regulated either by dilution of an inhibitor during cell growth (Pritchard *et al.*, 1969) or by accumulation of an autoregulated initiator (Sompayrac and Maaløe, 1973).

The DnaA initiator protein is considered to be the key element in this process as its transcription is both autoregulated (Braun *et al.*, 1985; Atlung *et al.*, 1985) and partially under stringent control (Atlung *et al.*, 1985; Chiaramello and Zyskind, 1989, 1990; Polaczek and Wright, 1990). Initiation is the major role of DnaA (Hirota *et al.*, 1970; Hansen and Rasmussen, 1977; Kaguni *et al.*, 1982) and in its absence the chromosome

can be replicated by an alternative (*oriC*-independent) pathway known as stable DNA replication (Kogoma and von Meyenberg, 1983; Kogoma *et al.*, 1985). DnaA overproduction has been shown to disrupt the cell cycle specificity of *oriC* minichromosomal replication (Pierucci *et al.*, 1989) even though the cellular concentration of DnaA did not vary appreciably over the cell cycle (Sakakibara and Yuasa, 1992). Løbner-Olesen and colleagues (1989) have shown convincingly, from studies where DnaA was under the control of the *lac* promoter in a *dnaA*^{ts} strain, that DnaA determines the initiation mass.

It is thought that the concentration of free, intracellular DnaA determines when initiation should occur (Hansen *et al.*, 1991). It has been calculated (Schaefer and Messer, 1991) that the chromosome contains roughly 1 600 DnaA boxes which could titrate DnaA away from *oriC*. DnaA has a low affinity for binding sites involved in *oriC* complex formation (Skarstad and Boye, 1994). Therefore, only at a certain critical DnaA concentration is there enough bound at the origin to trigger initiation. This is consistent with DNaseI-footprinting studies which showed that the R3 DnaA box is unprotected for most of the cell cycle (Samit *et al.*, 1989). It was proposed that DnaA binding to this DnaA box triggers initiation (Samit *et al.*, 1989).

Studies using *oriC* plasmids fully methylated at 5′-GATC sites showed them to be poorly transformed into *dam* strains and the reverse was found with unmethylated plasmids (Russel and Zinder, 1987). It was also found that unmethylated plasmids accumulated in a hemimethylated state, suggesting a block in the replication of hemimethylated plasmids. Ogden and colleagues (1988) have shown that the block in replication is due to specific binding of hemimethylated DNA to the outer membrane. Outer membrane preparations were shown to inhibit replication of hemimethylated DNA in crude solutions (Landoulsi *et al.*, 1990). However, DnaA binding at *oriC* before addition of the membrane preparation obviated this inhibition, suggesting that binding to the membrane prevents DnaA binding to the newly-synthesised daughter

strand. This mechanism of ensuring that the newly-replicated strand cannot be reinitiated has been termed sequestration (Russel and Zinder, 1987; Ogden *et al.*, 1988; Campbell and Kleckner, 1990; Landoulsi *et al.*, 1990; Hansen *et al.*, 1991; *Lu et al.*, 1994). Campbell and Kleckner (1990) also showed that the time required to convert hemimethylated to fully methylated DNA sites in both the *oriC* and *dnaA* promoter region was on average 4-fold longer than at any other region of the chromosome and this represented 30-40% of the cell cycle. Recently, the SeqA protein has been identified and found to act as a negative modulator of initiation by antagonising open complex formation, and to play a major role in sequestration of both *oriC* and the *dnaA* promoter regions (Lu *et al.*, 1994; Slater *et al.*, 1995). SeqA preferentially binds *oriC* that is hemimethylated, at the left hand side in the region of the 13-mer repeats (which start with a 5´-GATC). This blocks the action of Dam methyltransferase and prevents the daughter strand from being reinitiated (Lu *et al.*, 1994; Slater *et al.*, 1995).

1.2 Single-stranded Bacteriophage DNA Replication

The M13 (filamentous phage) and ϕX174 (icosahedral phage) genomes are quite similar in size, structure and organisation, and are likely to be linked evolutionarily. However, there are many biological differences in host range, virulence and genome replication.

Replication of the single-stranded DNA bacteriophages can be divided into three stages and depends on many of the same enzymes used in host chromosomal and plasmid replication (Figure 1.7). In the first stage (ss→RF), single-stranded viral DNA (ss) is converted to double-stranded replicative form (RF). No phage encoded proteins are required at this stage; the phage relies on the host replication system. In the second stage (RF→RF), amplification of replicative form DNA provides a template to support adequate levels of transcription. This step is initiated by the phage-encoded replication

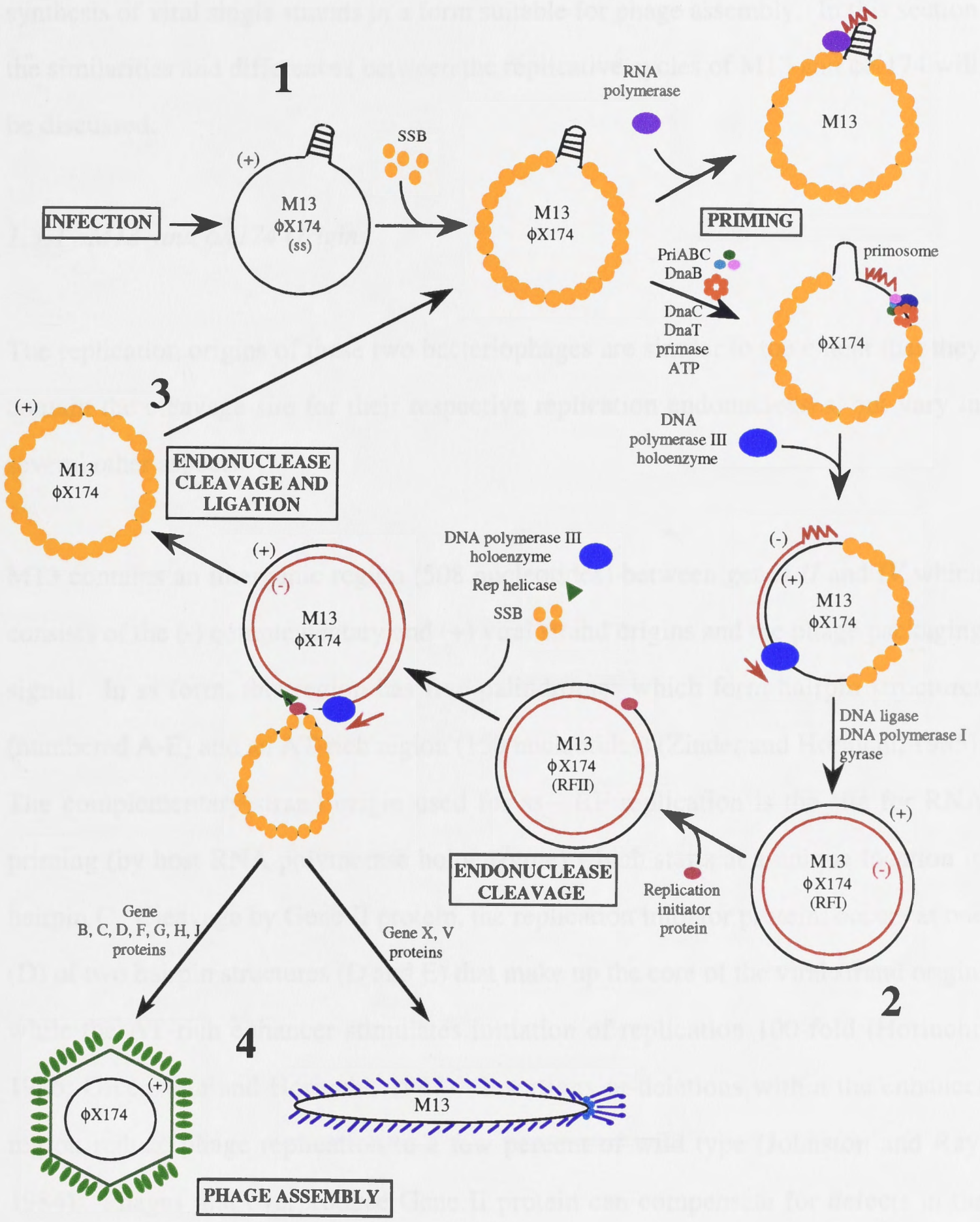


Figure 1.7

A schematic representation of replication of single-stranded DNA bacteriophages M13 and ϕX174. The similarities and differences between their genome replication are shown. The three stages of bacteriophage replication are ss→RF (1→2), RF→RF (2→3→2) and RF→ss (2→4).

initiator proteins, which are endonucleases. The final stage (RF→ss), involves synthesis of viral single strands in a form suitable for phage assembly. In this section the similarities and differences between the replicative cycles of M13 and ϕX174 will be discussed.

1.3.1 M13 and ϕX174 Origins

The replication origins of these two bacteriophages are similar to the extent that they contain the cleavage site for their respective replication endonucleases, but vary in several other aspects.

M13 contains an intergenic region (508 nucleotides) between genes *II* and *IV* which consists of the (-) complementary and (+) viral strand origins and the phage packaging signal. In ss form, this region has five palindromes which form hairpin structures (numbered A-E) and an AT-rich region (150 nucleotides) (Zinder and Horiuchi, 1985). The complementary strand origin used for ss→RF replication is the site for RNA priming (by host RNA polymerase holoenzyme) which starts at a unique location in hairpin C. Cleavage by Gene II protein, the replication initiator protein, occurs at one (D) of two hairpin structures (D and E) that make up the core of the viral strand origin, while the AT-rich enhancer stimulates initiation of replication 100-fold (Horiuchi, 1986; Greenstein and Horiuchi, 1987). Insertions or deletions within the enhancer region reduce phage replication to a few percent of wild type (Johnston and Ray, 1984). Phages that overproduce Gene II protein can compensate for defects in the enhancer but not for mutations in the core, indicating Gene II interaction occurs with the enhancer as well as the core (Dotto and Zinder, 1984a; Dotto and Zinder, 1984b). The *E. coli* integration host factor (IHF) also binds to the enhancer and is involved in many processes such as site specific recombination, plasmid replication and regulation of gene expression (Greenstein *et al.*, 1988). The DNA packaging signal (palindrome A) is important for sustaining a high yield of phage particles but is not required for

DNA replication (Dotto and Zinder, 1983).

In contrast to M13, the (+) and (-) strand origins ϕX174 (5.4 kb) are separated by approximately 2.5 kb. The viral (+) strand origin is a 30-bp region which is located within *gene A* and contains an AT-rich region and the recognition, binding and cleavage sites for the replication endonuclease Gene A (Baas *et al.*, 1980). The complementary strand origin is 55 nucleotides in length and forms a secondary structure which is the site for primosome assembly (denoted a *pas*, primosome assembly site, see Section 1.1.3 for details of the priming mechanism).

1.2.2 Stage 1: ss→RF

The first step in ss→RF replication is priming of the single-stranded viral DNA. M13 and ϕX174 employ different priming methods although both rely entirely on host proteins. The M13 viral, single stranded circle is primed by RNA polymerase holoenzyme containing σ^{70} which recognises a specific sequence on the (+) strand origin in the presence of SSB and produces a short RNA segment. In the case of ϕX174, RNA priming occurs at a primosome assembly site (*pas*) via a primosome consisting of DnaB, DnaG, PriA, PriB and possibly PriC proteins (see Section 1.1.3). In each case the RNA primer is extended by DNA polymerase III holoenzyme to form the product RFII (a duplex circle with a small gap or nick). The product contains an intact viral strand and a nearly full-length synthetic complementary strand with RNA primer covalently linked to the 5′ end. The RNA primer is removed by the 5′→ 3′ exonuclease action of DNA polymerase I. The *E. coli* proteins DNA ligase and gyrase seal and supercoil the product forming RFI (covalently-closed, circular, supercoiled DNA).

1.2.3 Stage 2: RF→RF

Specific cleavage of the supercoiled parental RF initiates synthesis and replication proceeds by a rolling-circle mechanism. The cleavage is performed by the phage encoded Gene II (M13) or Gene A (ϕX174) replicator initiator endonuclease, which is specific for the viral (+) strand origin of supercoiled RF. Rep helicase then forms a complex with the endonuclease and then proceeds to unwind the duplex DNA. The $3'\text{-}^{OH}$ terminus acts as a primer, elongated by DNA polymerase III holoenzyme in a rolling-circle manner as Rep helicase unwinds the duplex to allow *E. coli* SSB binding to the displaced viral strand.

A second endonuclease cleavage occurs upon recreation of a duplex origin and this is coupled to the ligation of the nascent strand into a single-stranded DNA circle. The endonuclease requires no energy co-factor suggesting that energy from cleavage is stored for subsequent ligation. The mechanism is intriguing but not yet understood. The newly-synthesised ssDNA circle is a template for complementary strand synthesis (ss→RF) by the same pathway that infecting viral ssDNA are converted to replicative form.

The endonucleases (Gene II and Gene A) act in an analogous manner to initiate DNA synthesis by cleavage of replication intermediates. Their genes also encode for a second protein (Gene X and Gene A*) that is translated from an internal in-frame AUG and are identical to the C-terminus of Gene II and Gene A, respectively. However, the two replication initiator proteins differ in that Gene A (ϕX174) forms a covalent linkage with the $5'\text{-}^{P}$ terminus upon cleavage (see Chapter 6); Gene II (M13) does not. It is thought that Gene II protein does remain associated with the $5'\text{-}^{P}$ terminus during replication. More information on these proteins is presented in Chapters 5 and 6.

1.2.4 Stage 3: RF→ss

Viral strand synthesis for packaging into phage proceeds by the same mechanism as RF→RF synthesis, in which cleavage initiates rolling-circle replication. At some point complementary strand synthesis is blocked and newly synthesised viral strands interact with the appropriate precursor proteins required for packaging.

For M13, the relative levels of phage-encoded Gene II, Gene X and Gene V proteins determine the switch between continuation of RF→RF synthesis and viral strand synthesis (Fulford and Model, 1988a, 1988b). When Gene V protein predominates, replication switches to viral strand synthesis. Gene X is thought to be involved in this switch by inhibiting complementary strand synthesis (Fulford and Model, 1984). Repression of Gene II and Gene X translation by Gene V may also play a role in regulation of the switch between replication pathways.

A different scenario occurs in ϕX174, where accumulated viral strands are coupled to morphogenesis of the phage particle and eight phage-encoded proteins are required (Kornberg, 1978; Meyer *et al.*, 1978; Eisenberg *et al.*, 1978). These proteins act to condense and encapsidate the single-stranded viral DNA, making it unavailable for conversion to RF. The accumulation of proheads (capsid precursors that assemble from Gene B, D, F, G and H proteins), Gene C and Gene J proteins causes the switch from RF→RF replication to viral strand synthesis and phage assembly. This is accomplished by Gene C protein binding to the RFII-Gene A-Rep complex as it unwinds the DNA, competing with SSB for the single-stranded DNA and inhibiting DNA synthesis. Proheads then recognise the complex containing Gene C, relieving the inhibition of DNA replication and allowing synthesis of viral ssDNA which is packaged directly.

1.3 Aims of this Thesis

Investigation of the several replicator initiator proteins from *E. coli* (DnaA) and its bacteriophages, M13 (Gene II protein) and ϕX174 (Gene A protein), was the focus of work presented in this thesis. These proteins have been extensively studied and they have been assigned a number of biochemical functions. However, the manner in which they perform some of these functions is unclear. There is also very little information known about the structure of these proteins. Therefore, our aim was to further characterise these proteins and, if possible, to obtain structural information by X-ray crystallography.

However, these proteins have proved to be difficult to obtain in large quantities and the methods of purification were laborious and returned poor yields. For example, both Gene II and DnaA proteins have a tendency towards aggregation or insolubility. Gene A represents a more extreme case in that its gene has not been able to be cloned and as a result protein is obtained only from infecting cells with ϕX174. Obtaining protein by this method is much less efficient than by the use of expression vectors. Thus, to obtain large quantities of these proteins for structural and functional characterisation, it was necessary to improve their levels of expression through the use of expression vectors and then develop new strategies for their large-scale purification. Some of that work is described.

CHAPTER 2

GENERAL MATERIALS AND METHODS

2.1 Bacterial Strains and Plasmids

The *Escherichia coli* strains used throughout this study and their genotypes are listed in Table 2.1. The strains prepared for overproduction of DnaA, Gene II and Gene A proteins and a DnaA mutant proteins are described in the text. Physical maps of the plasmid vectors used and their derivatives are presented in Figures 2.1 and 2.2.

2.2 Growth of Bacteria

Strains of *E. coli* were grown in LB medium (Luria and Burrous, 1957) supplemented with 25 μg/mL thymine (LBT) and, as required, with ampicillin (50 μg/mL) and/or chloramphenicol (30 μg/mL was used to maintain pLysS in host strains). Strains containing plasmids which direct protein expression from phage λ promoters were grown at 30°C, while others were grown at 37°C. For large-scale preparations of plasmid DNA, strains were grown in 56-minimal medium (Monod *et al.*, 1951) containing 1 mM magnesium sulfate and supplemented with caseine hydrolysate (20 mg/mL), trace salts (Gibson *et al.*, 1977), 40 mM glucose, 1 mg/mL vitamin B_1 and the appropriate antibiotic.

2.3 Chemicals, Reagents, Enzymes and Instruments

All chemicals and reagents used in this research were analytical grade and were obtained from commercial suppliers (Amersham, Amrad-Pharmacia, Ajax Chemicals, BDH Chemicals, BioRad, Boehringer-Mannheim, Cabot, Calbiochem, Difco, Fluka, Mallinckrodt Chemical Works, May and Baker, Qiagen, Sigma Chemical Company, Upjohn Company). Restriction endonucleases were obtained from Amersham International, Boehringer-Mannheim or New England Biolabs. Klenow enzyme was supplied by Boehringer-Mannheim, T4 DNA ligase by Bresatech. Lysozyme was purchased from Sigma Chemical Company. Calf intestinal alkaline phosphatase was

Table 2.1

Bacterial strains used in the experimental work.

Strain	**Genotype**	**Reference**
AN1459	*K12 supE44 thi-1 leuB6 thr-1 ilvC hsdR recA srlA::*Tn*10*	Elvin *et al.* (1986)
AN2666 (JM101*recA*)	*K12 supE44 thi-1 Δ(lac-proAB) recA srlA::*Tn*10*	Yanisch-Perron *et al.* (1985)
BL21(DE3)	*hsdS gal* (λ*c*I^{ts}857 *ind*1 *S*am7 *nin*5 *lac*UV5-T7 gene*1*)	Studier and Moffatt (1986)

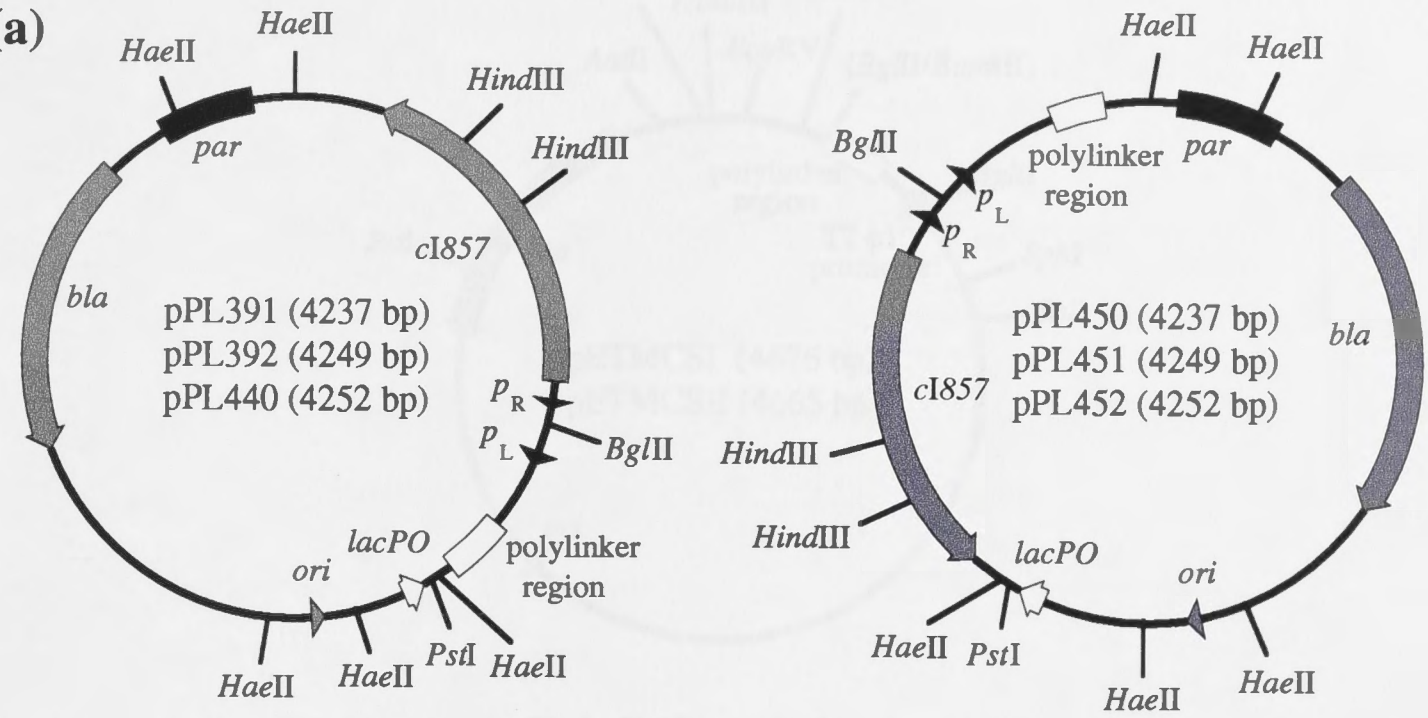
Figure 2.1

(a) The vectors pPL391, pPL392 and pPL440, although similar to pCE30 (Elvin *et al.*, 1990) are based on pMTL23P (Chambers *et al.*, 1988), a vector similar to the pUC plasmids. These vectors also contain the par region of pSC101. This region contributes greater stability at high copy number in the absence of ampicillin. These vectors also contain the *cI857*-p_R-p_L region as in pCE30 (Elvin *et al.*, 1990). pPL391 is an analogue of pCE30 (Elvin *et al.*, 1990) with a more extensive range of restriction endonuclease sites. The vector pPL392 is the corresponding analogue of pND201 (Elvin *et al.*, 1990) containing a strong RBS upstream of a unique *Hpa*I site. pPL440 contains the RBS of pPL392 eight base-pairs upstream of a potential ATG start codon as part of an *Nde*I restriction endonuclease site. Genes containing either an *Nde*I site at their start codon or which have been engineered to have one, may be inserted directly into pPL440 to achieve high-level expression.

pPL450, pPL451 and pPL452 are analogues of pPL391, pPL392 and pPL440 respectively. The vectors were constructed from pMTL22P and still contain the par region from pSC101. However, the *cI857*-p_R-p_L-polylinker region is in the reverse orientation.

(b) The phagmid pMA200 (Elvin *et al.*, 1990) is derived from and contains all the features of pCE30, with the addition of the viral strand origin region of filamentous phage f1. pMA200 derivatives, transformed into an M13-sensitive strain, replicate in the presence of superinfecting M13KO7 to produce packaged single-stranded plasmid DNA. If the gene is in the correct orientation for expression from p_R-p_L, single-stranded plasmid DNA contains the coding strand. The packaged single-stranded DNA isolated from the culture supernatant can be used for DNA sequencing or modification by site-directed mutagenesis. On transfection, mutant plasmids are then immediately available for high-level expression of the mutant protein.

(a)



Polylinker regions:

pPL391 & pPL450: ..AGAAGGGCAGCATTCAAAGCAGAAGGCTTTGGGGTGTGTGATACGAAAC

Primer 9 → ← Primer 1

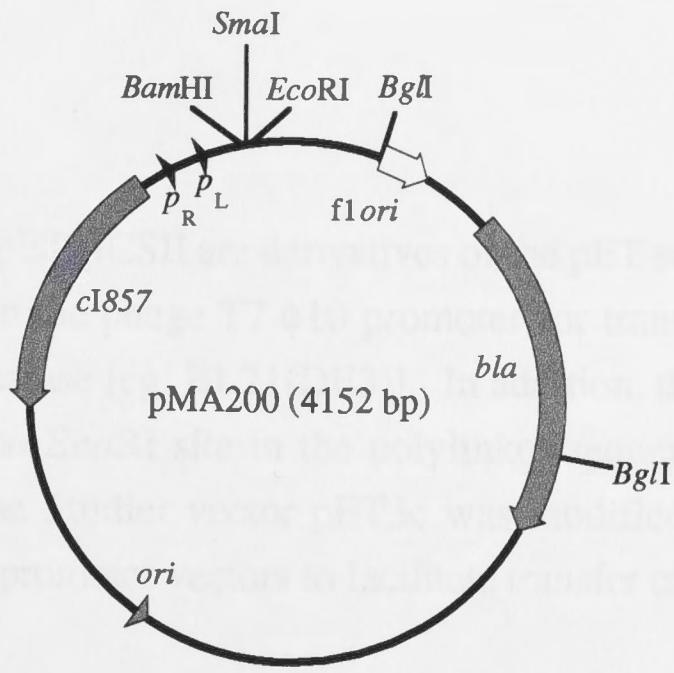
BamHI AatII MluI PstI HindIII EcoRI
GAAGCATTGGGATCCATATGACGTCGACGCGTCTGCAGAAGCTTCTAGAATTC
NdeI SalI/AccI XbaI

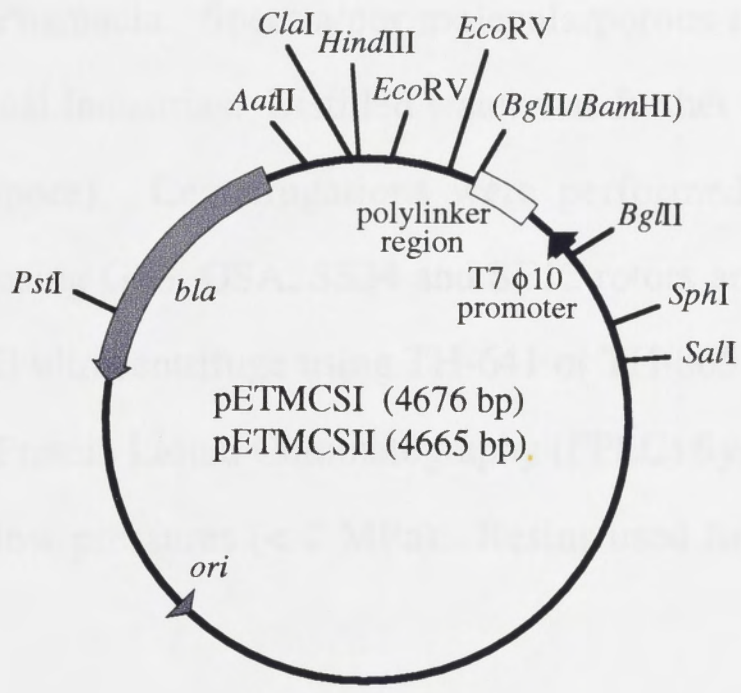
SstI SmaI NcoI ClaI BglII StuI
GAGCTCCCGGGTACCATGGCATGCATCGATAGATCTCGAGGCCTCGCGAGCT..
KpnI SphI XhoI NruI

pPL392 & pPL451: ..GGATCCTAAGGAGGTTAACTATGACGTC..
BamHI HpaI AatII
RBS

pPL440 & pPL452: ..GGATCCTAAGGAGGTTGATCCATATGACGTC..
BamHI AatII
RBS NdeI

(b)





Polylinker regions:

Figure 2.2

Vectors pETMCSI and pETMCSII are derivatives of the pET series of vectors (Studier *et al.*, 1990) and contain the phage T7 ϕ10 promoter for transcription in strains that express T7 RNA polymerase [eg: BL21(DE3)]. In addition, the *Eco*RI site of pET3c was removed so that the *Eco*RI site in the polylinker sequence could be used as a useful cloning site. The Studier vector pET3c was modified to contain polylinker regions similar to the λ-promoter vectors to facilitate transfer of genes as casettes from one system to the other.

supplied by Amrad-Phamacia. Spectra/por molecularporous membrane tubing was from Spectrum Medical Industries. Distilled water was further purified with a milliQ (mQ) system (Millipore). Centrifugations were performed in a Sorvall RC5C centrifuge (Dupont) using GS3, GSA, SS34 and SE12 rotors and ultracentrifugations in a Sorvall OTD 75B ultracentrifuge using TH-641 or TH-865 rotors. Proteins were purified using a Fast Protein Liquid Chromatography (FPLC) System (Pharmacia) with pumps operating at low pressures (< 2 MPa). Resins used for chromatography are described in the text.

2.4 Preparation of Plasmid DNA

2.4.1 *Plasmid Extraction by Alkaline Lysis*

Small-scale preparations of plasmid DNA were performed essentially as described (Silhavy *et al.*, 1984) from cells grown on LBT plates containing the appropriate antibiotics (Section 2.2). Plasmid DNA obtained from this procedure was used for analytical purposes or transformation of competent bacterial cells.

2.4.2 *Large-scale Plasmid Preparation*

Large-scale preparations of highly-purified plasmid DNA was obtained by two methods: preparation using the commercially available kits (QIAGEN) or centrifugation to equilibrium in CsCl gradients.

Plasmid purification by the QIAGEN maxi (up to 500 µg) preparation was carried out using protocols supplied by the manufacturer (Qiagen, Inc.).

Alternatively, large-scale isolations of highly-purified plasmid DNA were by two successive bandings in CsCl density gradients from cells grown in minimal media.

Plasmid DNA was amplified with spectinomycin. The cells were lysed with Triton X-100 using a modification of the procedure described by Davis *et al.* (1980). All procedures were carried out at 0°C unless otherwise indicated.

An overnight culture of the *E. coli* strain containing the plasmid was used to inoculate 1 L of 56-minimal medium containing the particular antibiotic and was then aerated at 30°C. When the absorbance of the culture reached A_{595} = 0.5, spectinomycin was added (300 mg/L), and areation continued at 30°C for a further ~16 h. The cells were harvested by centrifugation (8 000 x g, 10 min, 4°C) and resuspended in 6.25 mL of ice-cold 50 mM Tris.HCl pH 8.0, 25% (w/v) sucrose, frozen in liquid nitrogen and stored at -70°C until required.

The cell suspension was thawed, diluted with 1.25 mL of resuspension buffer containing lysozyme (10 mg/mL) and swirled for 5 min. Then, 1.25 mL of 500 mM EDTA (pH 8.5) was added and the flask swirled for a further 5 min. Following the method of Katz *et al.* (1973), cell lysis was performed with the addition of Triton X-100 solution (10 mL; 0.1% v/v Triton X-100, 50 mM Tris.HCl pH 8.0, 62.5 mM EDTA). After 10 min, the lysate was centrifuged (48 000 x *g*, 1 h, 4°C). The supernatant was collected and its density adjusted by dilution to 25 mL with mQ H_2O and addition of 24.38 g CsCl. After centrifugation (13 000 x *g*, 1 h, 4°C), the supernatant was decanted through a tissue filter into a Sorvall T-865 polyallomer tube. The tube was filled with 2.55 mL of a solution of 10 mg/mL ethidium bromide and TE buffer (10 mM Tris.HCl pH 7.4, 1 mM EDTA) to which 0.975 g/mL CsCl had been added. The plasmid DNA was isolated in the density gradient produced by centrifugation (125 000 x *g*, > 40 h). The plasmid DNA band, visualized using a long-wave UV lamp, was collected from tubes using an 18-gauge hypodermic needle and syringe as described (Sambrook *et al.*, 1989). The plasmid DNA was transferred to a T-1270 polyallomer tube, filled with 0.85 mL of 10 mg/mL ethidium bromide solution and TE buffer containing CsCl (0.975 g/mL). Following centrifugation (125 000 x *g*,

> 40 h), the plasmid DNA was isolated as before. The ethidium bromide was removed by repeated extractions with an equal volume of propan-2-ol saturated with 5 M NaCl in TE buffer. The solution containing plasmid DNA was dialysed against 2 changes of 1 L of TE buffer over 24 h. The concentration of ds DNA was determined spectrophotometrically assuming a solution with A_{260} = 1 contained 50 µg/mL DNA. Plasmid DNAs were routinely stored in TE buffer at -20°C.

2.5 Restriction-endonuclease Digestion of DNA

Restriction-endonuclease digests were performed in buffers supplied by the manufacturers (Section 2.3). When two or more restriction enzymes with different requirements of salt concentration were used, the digest requiring lower salt concentration was performed first. The concentration of salt was elevated for digestion with the second enzyme. In most cases, restriction-endonuclease digestions were carried out at 37°C for 1 h and were terminated by addition of 0.5 volumes of restriction-endonuclease stop mix (50 mM EDTA, 17% v/v glycerol, 0.07% bromophenol blue, pH 8.5) at 0°C.

For partial restriction-endonuclease digestion, the reaction was optimized on a small-scale before the plasmid DNA was digested on a preparative scale. A progressive series of two-fold dilutions of reaction mixtures containing identical amounts of plasmid DNA were prepared. After a 30-min treatment at 37°C, the reactions were stopped by addition of restriction-endonuclease stop mix at 0°C. Electrophoresis on agarose gels (Section 2.7) was used to evaluate the results of the series of digests. The optimal conditions were determined from analysis of the fragment patterns the optimal conditions were determined and used in a scaled-up partial digestion.

2.6 DNA End-filling with the Klenow Enzyme

Recessed 3´-termini created by restriction-endonuclease digestion were filled-in using the large fragment of DNA polymerase I (Klenow enzyme) as described by Sambrook *et al.* (1989).

2.7 Electrophoresis of DNA

Agarose gels for electrophoresis of DNA were cast in a Davis system horizontal submarine apparatus (Sambrook *et al.*, 1989) using a toothed comb to form wells for loading samples. Gels (147 x 136 x 10 mm) contained agarose at concentrations in the range 0.7 to 2% (w/v) in either TBE (89 mM Tris-borate, 2 mM EDTA; Sambrook *et al.*, 1989) or TAE (40 mM Tris-acetate, 1 mM EDTA; Sambrook *et al.*, 1989) containing 0.5 µg/mL ethidium bromide. DNA samples were mixed with a 1/4 volume of loading buffer (2% bromophenol blue, 50% glycerol). Samples were loaded onto a gel and electrophoresed at 40 V using a Bio-Rad 2000/02 power supply until the DNA had entered the gel matrix. Electrophoresis was then continued at 100 V until the required resolution of DNA fragments was achieved.

Fragments of DNA in agarose gels separated by electrophoresis were visualized using a long-wave UV lamp or a short-wave transilluminator (UV transilluminator Model TS-15, Ultraviolet Products). Photographs of gels placed on the transilluminator were made using a model MP-4 Land Camera (Polaroid) with Polaroid film (114 x 89 mm type 57/high-speed). λ *cI857Sam7* DNA (Boehringer-Mannheim) which had been digested with *Eco*RI and *Hind*III restriction enzymes was used as a DNA fragment size standard.

2.8 Preparation of Oligonucleotides

Synthetic oligonucleotides used in cloning and site-directed mutagenesis experiments were either purchased from Auspep or prepared at the Biomolecular Resource Facility (Centre for Molecular Structure and Function, ANU) using an Applied Biosystems 380B DNA Synthesizer. The oligonucleotides were deprotected following a 12 h treatment at 56°C in ammonia solution, and dried. Before use, the dried oligonucleotides were resuspended in 500 µL of TE buffer and the solution clarified in a microcentrifuge (MSE Micro Centaur Microfuge) at 4°C for 10 min. The concentrations of the oligonucleotides were determined by measurements of A_{260}. Molar extinction coefficients of the particular oligonucleotides were calculated from their nucleotide composition and molar extinction coefficients of nucleotides at 260 nm (Sambrook *et al.*, 1989).

2.9 Isolation and Purification of DNA Fragments

Electroelution of DNA fragments from agarose gel onto NA45 membrane (Schleicher and Schuell) or extraction of DNA from a gel slice using the QIAEX Gel Extraction Kit (QIAGEN) were methods used for recovery of DNA fragments that had been separated by agarose gel electrophoresis. These methods were generally used for isolation of fragments generated by restriction-endonuclease digestion and for the removal of enzymes after end-filling with the Klenow enzyme and dephosphorylation.

The methods were carried out as described by Sambrook *et al.* (1989) or by protocols provided by the manufacturer (QIAGEN). The DNA recovered from gels was further purified by phenol/chloroform extraction (Sambrook *et al.*, 1989). Extractions with phenol and chloroform were also used for removal of enzymes following end-filling with the Klenow enzyme and dephosphorylation.

2.10 5´-Dephosphorylation of Linear Plasmid DNA

To suppress self-ligation of vector DNA, the 5´-phosphate groups of plasmid DNA linearized by restriction endonucleases were removed. Dephosphorylation of the linear plasmid DNA was carried out essentially as described by Sambrook *et al.* (1989) using calf intestinal alkaline phosphatase.

2.11 Ligation

Ligations were carried out with T4 DNA ligase and purified fragments of DNA (Section 2.9). Concentrations of insert and vector fragments in a particular ligation reaction were adjusted according to recommendations of Legerski and Robberson (1985), unless otherwise stated.

2.11.1 Ligation of Cohesive Termini

DNA fragments with cohesive termini were ligated using ligation buffer described by Sambrook *et al.* (1989). Reactions were routinely carried out at 14°C for 16 h in a volume of 40 µL.

2.11.2 Ligation of Blunt-ended Termini

Fragments of DNA with blunt-ended termini were ligated in blunt-end ligation buffer (Sambrook *et al.*, 1989) at 30°C for 2 h in volumes of 20 to 50 µL.

2.12 Dideoxy Sequencing of DNA

Determination of the nucleotide sequences of plasmid DNA was carried out using either the ABI PRISM dye terminator cycle sequencing ready reaction kit with Ampli Taq

DNA polymerase FS (Perkin Elmer) or the dideoxy-mediated chain termination reaction (Sanger *et al.*, 1977).

The ABI PRISM method (Perkin Elmer) was carried out according to the manufacturer's specifications and DNA sequences were analysed on the ABI 373A DNA Sequencer.

The procedure for DNA sequencing by the dideoxy-mediated chain termination reaction was essentially as described by Tabor and Richardson (1987) using T7 DNA polymerase, with [α-^{32}P]dATP (~ 1.1 x 10^{11} Bq/mmol, Amersham) for detection. Primers for sequencing of pPL450, pPL451, pPL452 and their derivatives are shown in Figure 2.1. Other primers used are described in the text.

Analysis of sequencing reactions was carried out on 6% w/v polyacrylamide gels (19:1 acrylamide:bisacrylamide) in TBE buffer containing 8 M urea. Slab gels (500 x 380 x 0.4 mm) were mounted on a Bio-Rad Sequi-gen DNA nucleic acid sequencing apparatus containing TBE electrophoresis buffer and pre-electrophoresed for 30 min at 50-55°C and 1500 V (Bio-Rad model 3000Xi electrophoresis power supply). Samples in formamide loading buffer were treated at 95°C for 2 min, then loaded and electrophoresed. Following electrophoresis, gels were fixed to 3MM filter paper, covered with plastic film and dried (Bio-Rad Model 483 Slab Drier).

Dried polyacrylamide gels containing [^{32}P]-labelled DNA were autoradiographed with XAR5 film (Kodak) in a light-proof cassette (Kodak) at ambient temperature for 6-12 h. Films were developed on a Kodak X-Omatic automatic X-ray film developer.

2.13 Transformation of *E.coli*

2.13.1 Preparation of Competent Cells

Cells of *E. coli* were made competent for transformation with plasmid DNA using $CaCl_2$ (Morrison, 1979). Cells were either used immediately or stored for up to 6 months at -70°C in 15% glycerol.

2.13.2 Transformation of Competent Cells

Competent cells were transformed with plasmid DNA using the method of Morrison (1979), except that after addition of plasmid DNA and 30 min storage at 0°C, they were routinely treated at 30°C for 2 min before incubation in LBT broth. LBT plates containing the appropriate antibiotic were used for selection of transformants.

2.14 Determination of Protein Concentration

Concentrations of proteins were determined by either the protein-dye binding method of Bradford (1976) or by A_{280} measurement. Bradford reagent was purchased from Bio-Rad; bovine serum albumin (BSA; Pharmacia) was used as a standard. A_{280} measurements were carried out using either a Varian Cary 1E UV-vis Spectrophotometer or a Hewlett-Packard 850A UV-vis Spectrophotometer in 1-mL quartz cuvettes. Molar extinction coefficients of proteins were estimated from their amino acid composition as described by Gill and von Hippel (1989).

2.15 SDS - Polyacrylamide Gel Electrophoresis

Electrophoresis of proteins was performed under denaturing conditions as described by Laemmli (1970).

The resolving gel contained 12.5% (w/v) acrylamide (30:2.7 acrylamide:bisacrylamide), 375 mM Tris.HCl pH 8.8, 0.1% (w/v) SDS, 0.033% (w/v) ammonium persulfate and 0.033% (w/v) TEMED. Stacking gels contained 4.5% acrylamide (30:2.7 acrylamide:bisacrylamide), 125 mM Tris.HCl pH 6.8, 0.1% (w/v) SDS, 0.08% ammonium persulfate, 0.08% TEMED. Slab gels (1.5 x 200 x 150 mm) were prepared in vertical glass moulds. The buffer for SDS-PAGE contained 51 mM Tris base, 384 mM glycine and 0.1% (w/v) SDS. Protein samples were mixed with an equal volume of sample buffer [300 mM Tris base, 15% (v/v) glycerol, 0.6% (w/v) bromophenol blue, 50 mM DTT, 2% SDS] and heated at 95°C for 2 minutes.

BioRad 2000/02 or Pharmacia EPS 500/400 instruments were used as power supplies. A potential difference of 60 V was applied until samples reached the interface between the stacking and resolving gels. Resolution of proteins was achieved by further electrophoresis at 140 V.

Gels were fixed and stained for a minimum of 2 hours in staining solution [40% (v/v) methanol, 10% (v/v) acetic acid, 0.3% (w/v) Coomassie Brilliant Blue R] and destained in buffer containing 10% (v/v) propan-2-ol and 10% (v/v) acetic acid. Photographs for laboratory documentation were taken using a Polaroid camera (model MP-4 Land Camera) and Polaroid Type 55 high-speed film. Photographs presented in this thesis were taken by Mr B. Wight from the Photography Unit (Research School of Chemistry).

CHAPTER 3

OVERPRODUCTION, PURIFICATION AND CRYSTALLISATION OF THE DnaA PROTEIN

3.1 Aims and Significance

The DnaA protein is a key actor in initiation of DNA replication at the origin (*oriC*) of the *E. coli* chromosome. This complex protein has multiple functions which include sequence-specific DNA binding at *oriC* and in the regulatory regions of several genes, binding of ATP, ADP and cAMP, formation of a nucleoprotein complex at *oriC*, loading of the DnaB helicase onto ssDNA, and interaction with acidic phospholipids and other proteins (Funnel *et al.*, 1986; Sekimizu *et al.*, 1987; Bramhill and Kornberg, 1988; Sekimizu and Kornberg, 1988; Sekimizu *et al.*, 1988; Hughes *et al.*, 1988; Yung *et al.*, 1990). Although the protein has been characterised, the finer details of many of these processes are still to be determined. As generous quantities of DnaA are required for structural and spectrocopic studies to probe the protein's multiple functions, improvement in its overproduction and purification was necessary.

The DnaA protein has been overproduced in strain N4830 (*cI857*) containing pBF1509 (Fuller and Kornberg, 1983). However, high copy number λ promoter vectors developed in our laboratory have been shown to overproduce proteins to very high levels (e.g., see Elvin *et al.*, 1990, Love *et al.*, 1996). These vectors were used in an effort to increase the level of *dnaA* gene expression. Also, the nucleotide distance between the RBS and ATG start codon was examined to determine its effect on protein expression. The combination of these studies was used to improve DnaA overproduction such that large quantities could be obtained from only a few litres of cell culture.

The previously-reported methods for purification result in observation of aggregated (inactive) and monomeric (active) forms of DnaA (Fuller and Kornberg, 1983; Sekimizu *et al.*, 1988). It has been reported that 20-50% of overproduced DnaA aggregated upon purification and was insoluble at low ionic strength. Aggregation of DnaA was alleviated by treatment with guanidine.HCl and further purification relied on

the correct refolding of the protein on rapid removal of the denaturant. Our goal was to design a strategy to purify fully active, monomeric DnaA without the use of denaturants in a quantity sufficient for structural and functional charaterisation with an emphasis on structure determination by X-ray crystallography.

Recently, the C-terminal 94 amino acids of DnaA has been identified as the DNA-binding region and has been designated as Domain IV (Roth and Messer, 1995). We also aimed to clone, overproduce and purify Domain IV (DnaAC::94) for structure determination by X-ray crystallography or NMR.

3.2 Introduction

The *Escherichia coli* DnaA is a 52-kDa protein which plays a central role in the initiation of DNA replication of the *E. coli* chromosome at the origin of replication, *oriC*. It binds ATP (K_d = 0.03 μM) and very slowly hydrolyses it to ADP, although ATP hydrolysis is not required for initiation (Sekimizu *et al.*, 1987). Both the ADP- and ATP-bound forms are able to form a complex with *oriC* (Sekimizu *et al.*, 1987). However, the complexes differ and only the ATP-bound form is capable of initiation of DNA replication. Low affinity binding of ATP is also required and binding of cAMP (K_d = 1 μM) stimulates binding to *oriC* (Yung *et al.*, 1990; Hughes *et al.*, 1988). DnaA binds specifically to 9-bp DnaA boxes (5′-TTAT$^{C}/_{A}$CA$^{C}/_{A}$A) which are located in *oriC* and other replication origins (P1 *oriR*, F *oriS*, R1 and R6K *oriγ*) as well as in the control region of its own gene and of several others (eg., the *gua* operon, *uvrB*, *rpoH* and *mioC*) (Tesfa-Selase and Drabble, 1992; Vanden Berg *et al.*, 1985; Wang and Kaguni, 1987; Løbner-Olesen *et al.*, 1985). DnaA also interacts with the molecular chaperone DnaK (Malki *et al.*, 1991), and phospholipids (cardiolipin, phosphatidylethanolamine and phosphatidylglycerol), and interaction with the cell membrane has been proposed (Sekimizu *et al.*, 1988).

During initiation, DnaA binds cooperatively to four DnaA boxes in *oriC*, stimulating the formation of a nucleoprotein complex with the DNA wrapped around the periphery of the protein (Fuller *et al.*, 1984; Woelker and Messer, 1993). Approximately 10-30 monomers are thought to be involved, as judged by electron microscopy and binding studies (Funnel *et al.*, 1987). A temperature-dependent structural change in the DnaA-*oriC* complex results in localised melting of the duplex DNA within the 13-bp AT-rich repeats positioned at the leftmost edge of *oriC* (Bramhill and Kornberg, 1988). Once the strands have been separated, DnaA interacts with the DnaBC complex to load DnaB helicase onto the single-stranded DNA (Funnel *et al.*, 1986) (Figure 1.3).

The *dnaA* gene is the first of four genes in an operon (*dnaA-dnaN-recF-gyrB*) mapped to the 83 minute region of the *E. coli* chromosome (Miki *et al.*, 1978; Miki *et al.*, 1979) (Figure 3.1). Genes similar to the *dnaA* gene of *E. coli* have been identified throughout the bacterial world and in many cases the same gene order (*dnaA-dnaN-recF-gyrB*) was found (Skarstad and Boye, 1994). Transcription proceeds counter clockwise relative to the *E. coli* map from promoters located 240 (DnaAp1) and 160 (DnaAp2) nucleotides upstream of the translational start site (Figure 3.1) (Hansen *et al.*, 1982b). Two DnaA boxes, one between the promoters and one in the structural gene, and six Dam (deoxyadenosine methyltransferase) sites located in the promoter region are involved in the regulation of transcription of the operon (Braun *et al.*, 1985; Atlung *et al.*, 1985; Kucherer *et al.*, 1986; Braun and Wright, 1986; Campbell and Kleckner, 1990).

The DnaA protein consists of 467 amino acids which have been divided into four domains based on their function and sequence alignment (Figure 3.2) (Fujita *et al.*, 1992; Skarstad and Boye, 1994; Roth and Messer, 1995). Domain I (amino acids 1-53) is thought to be involved in DNA replication as polyclonal antibodies raised against the first 55 amino acids inhibited replication *in vitro* (Skarstad and Boye, 1994). Very little is known about the function of Domain II (amino acids 54-128) while Domain III

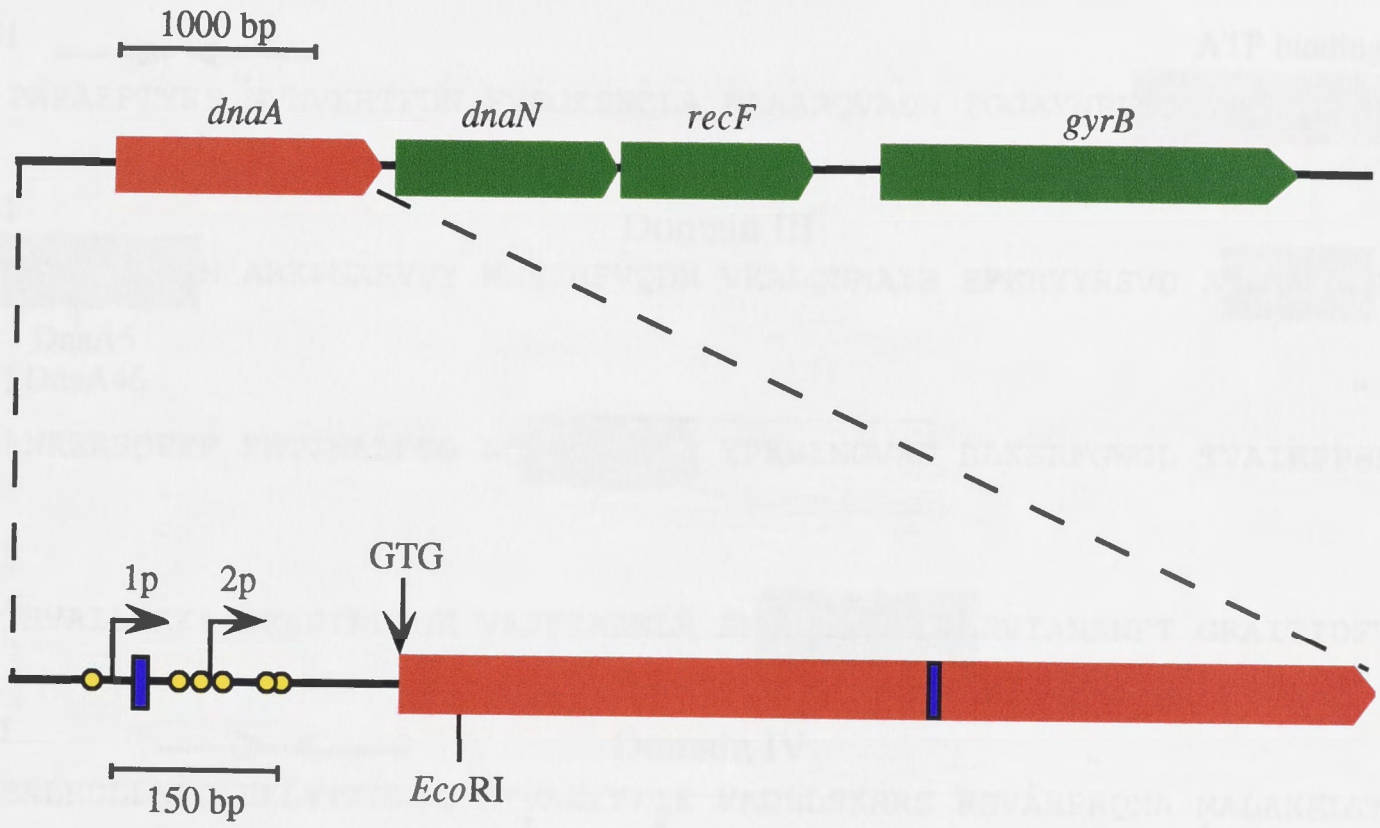


Figure 3.1

A schematic representation of the *dnaAN* operon. The *dnaA* gene is the first of four genes in the operon. The other genes are: *dnaN*, encoding the β-subunit of the DNA polymerase holoenzyme (Yuasa and Sakikabara, 1980; Sakikabara *et al.*, 1981); *recF*, a gene necessary for a DNA recombinational pathway (Ream *et al.*, 1980); and *gyrB*, a gene encoding one of the two subunits of DNA gyrase (Hansen and Von Meyenburg, 1979). The enlarged *dnaA* gene (bottom) shows the two *dnaA* promoters (black arrows), DnaA boxes (5´-TTAT$^{C}/_{A}$CA$^{C}/_{A}$A, blue boxes) and Dam methyltransferase sites (5´-GATC, yellow circles).

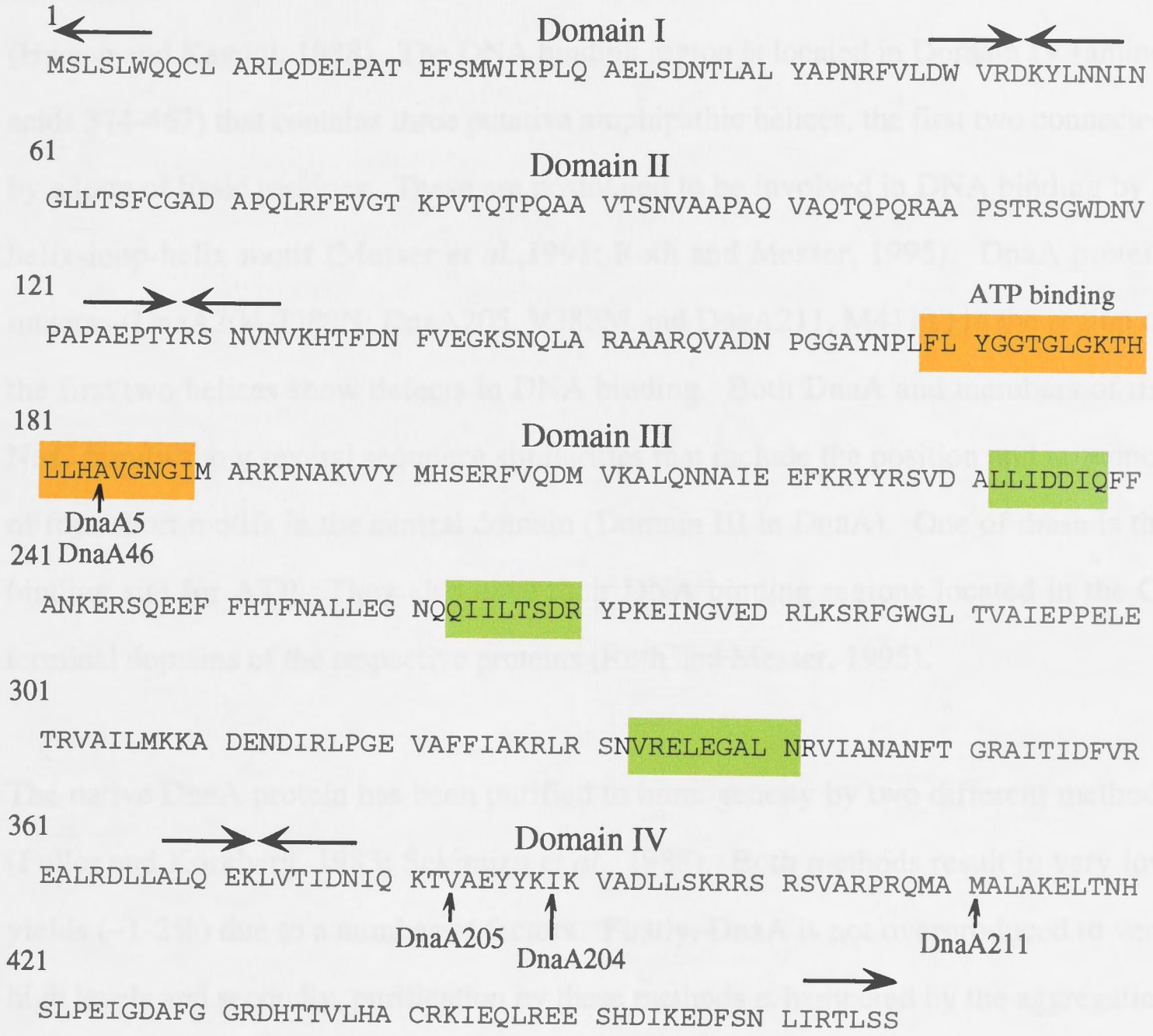


Figure 3.2

The amino acid sequence of the DnaA protein. The four domains of the protein are shown with black arrows indicating the junctions between the domains. The ATP binding region (yellow box) and three other regions (green boxes) which are highly conserved among DnaA homologues and members of the NtrC family are also shown. The positions of DnaA mutants DnaA5 and DnaA46 (A184V), DnaA205 (V383M), DnaA204 (I389N) and DnaA211 (M411T) are indicated.

(amino acids 129-373) harbours the ATP binding site (consensus sequence G-X-X-G-X-G-K-T) and three other conserved regions which are related to the NtrC family of transcription factors that also bind ATP (Koonin, 1993). The ATP binding site was identified from mutants DnaA5 and DnaA46 (A184V) which were unable to bind ATP (Hwang and Kaguni, 1988). The DNA binding region is located in Domain IV (amino acids 374-467) that contains three putative amphipathic helices, the first two connected by a loop of basic residues. These are postulated to be involved in DNA binding by a helix-loop-helix motif (Messer *et al.*,1991; Roth and Messer, 1995). DnaA protein mutants (DnaA204, I389N; DnaA205, V383M and DnaA211, M411T) in the region of the first two helices show defects in DNA binding. Both DnaA and members of the NtrC family show several sequence similarities that include the position and sequence of four short motifs in the central domain (Domain III in DnaA). One of these is the binding site for ATP. They also have their DNA binding regions located in the C-terminal domains of the respective proteins (Roth and Messer, 1995).

The native DnaA protein has been purified to homogeneity by two different methods (Fuller and Kornberg, 1983; Sekimizu *et al.*, 1988). Both methods result in very low yields (~1-2%) due to a number of factors. Firstly, DnaA is not overproduced to very high levels and secondly, purification by these methods is hampered by the aggregation of monomers (up to 50% of the DnaA) and irreversible precipitation of the protein at low ionic strength. The aggregated form of the protein was inactive in a reconstituted replication system (Fuller and Kornberg, 1983; Sekimizu *et al.*, 1988). As a result, large cultures were required to obtain milligram quantities of DnaA and the denaturant guanidine.HCl was employed to overcome monomer aggregation and low ionic strength precipitation. Large cultures make the purification laborious and tedious while purification subsequent to use of guanidine.HCl relies on correct refolding of the protein into an active conformation. Clearly there was a need to improve the overproduction of DnaA and to devise a new method of purification to obtain large quantities of highly purified, fully-active, monomeric protein for structural and

functional charaterisation. Conditions were required that could overcome the problems associated with protein aggregation and low ionic strength precipitation whilst avoiding the use of denaturants.

3.3 Materials and Methods

3.3.1 Overproduction of DnaA Protein

Strains RSC625 (AN1459/pPL400 and p179) and RSC668 (AN1459/pND556) were used to overproduce DnaA. Plasmid p179 (Zylicz *et al.*, 1983) directs overproduction of DnaK protein. Plasmids pPL400 and pND556 direct high level overproduction of DnaA protein and were constructed by P. E. Lilley and Dr N. E. Dixon (Figure 3.3). Both plasmids are derivatives of pPL391 (Figure 2.1) with the *dnaA* gene under the control of the strong bacteriophage λ promoters p_R and p_L. The natural GUG start codon in the *dnaA* mRNA was replaced with the more usual AUG in both plasmids since it is believed that GUG is less efficiently utilised by ribosomes (Singer *et al.*, 1981) (Figure 3.4). These plasmids also contain a synthetic RBS which is perfectly complementary to the 3′ end of *E. coli* 16S rRNA spaced optimally from the start of the *dnaA* gene (Figure 3.4). They differ in that the DnaA-encoding region of pPL400 was derived from pBF1509 (Fuller and Kornberg, 1983) while pSJS9 (Blanar *et al.*, 1984) was the source of most of *dnaA* for pND556. Essentially, both plasmids are otherwise identical except that pPL400 (from pBF1509) contains a small portion of pBR322 DNA at the end of the *dnaA* region. The plasmids also direct expression of the thermolabile λ *cI857* repressor. When strains containing pPL400 or pND556 are grown at 30°C, the repressor is active and prevents transcription from the tandem λ promoters. Upon elevation of the temperature to 42°C, the repressor is inactivated, enabling transcription.

The strain RSC625 (AN1459/pPL400 and p179) was used for co-overproduction of

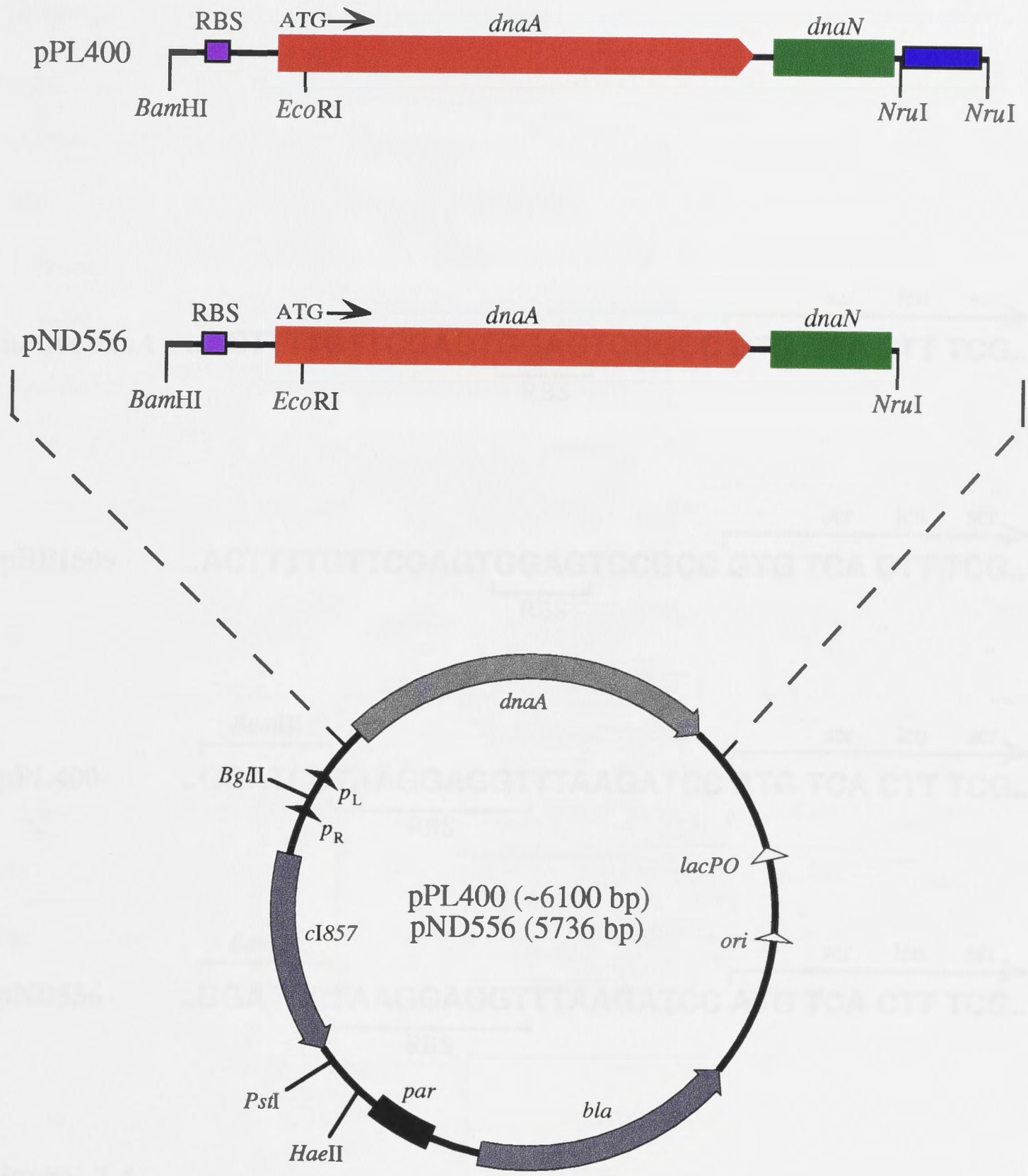


Figure 3.3

Plasmids pPL400 and pND556 that direct thermoinducible overproduction of DnaA. Both plasmids contain the same *dnaA* gene derived from different sources. pPL400 contains in addition small fragment of DNA from pBR322. pBF1509 (Fuller and Kornberg, 1983) provided the *dnaA* gene in pPL400 and pSJS9 (Blanar *et al.*,1984) was the source of *dnaA* in pND556. In both plasmids the GTG start codon has been changed to ATG downstream of a synthetic RBS (purple box) (5´-TAAGGAGGTT) which is perfectly complementary to the 3´ end of the *E. coli* 16S rRNA.

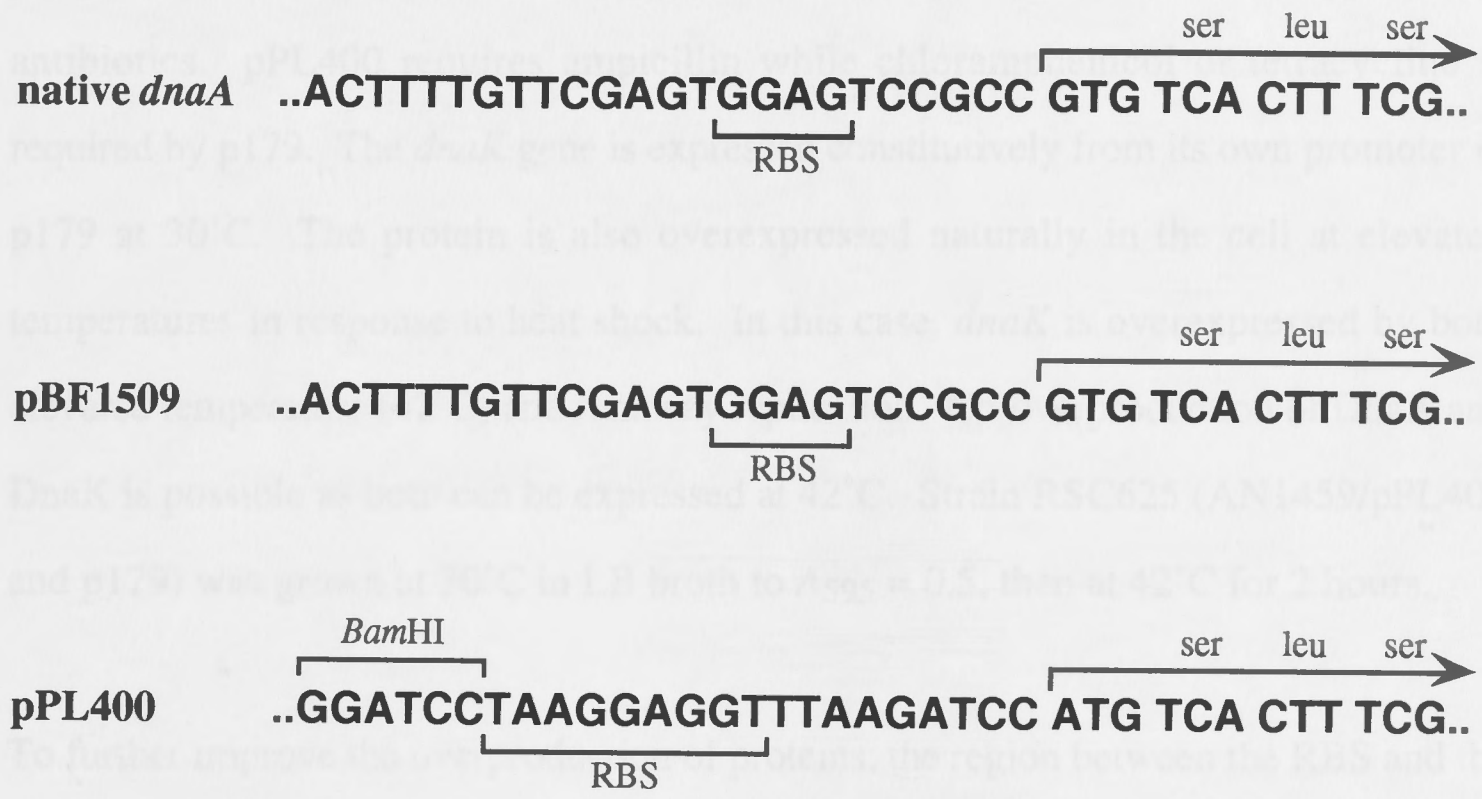


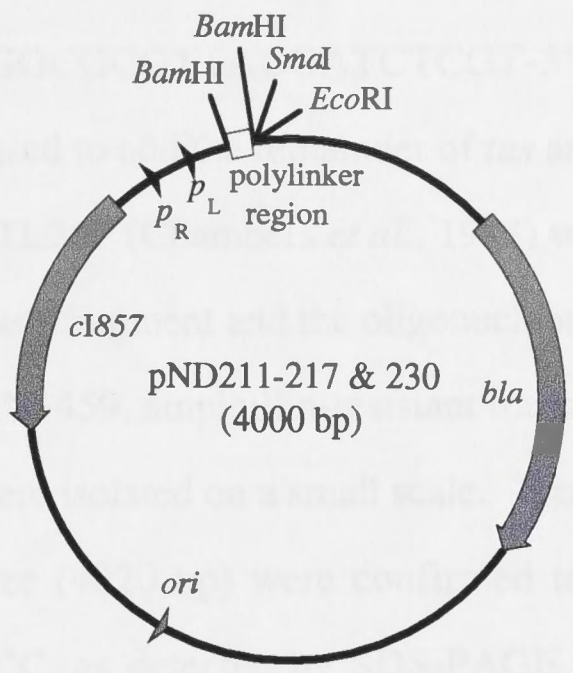
Figure 3.4

The native *dnaA* sequence and *dnaA* clones showing alterations in the start codon, RBS and the region between the two elements. The methionine residue at the start is cleaved *in vivo* and the amino acid sequence is not altered.

DnaA and DnaK proteins. This strain contained pPL400 (described above) and p179 (pMOB45 *dnaK*$^+$) (Zylicz *et al.*, 1983) for overproduction of DnaA and DnaK, respectively. pMOB45 (Bittner and Vapnek, 1981) is a vector derived from plasmid R1 that employs runaway replication for increased expression of cloned genes. Increased levels of protein are due to the presence of high levels of plasmid. Both plasmids are stably maintained together in the cell in the presence of appropriate antibiotics. pPL400 requires ampicillin while chloramphenicol or tetracycline is required by p179. The *dnaK* gene is expressed constitutively from its own promoter in p179 at 30°C. The protein is also overexpressed naturally in the cell at elevated temperatures in response to heat shock. In this case, *dnaK* is overexpressed by both elevated temperature (42°C) and runaway replication. Co-overproduction of DnaA and DnaK is possible as both can be expressed at 42°C. Strain RSC625 (AN1459/pPL400 and p179) was grown at 30°C in LB broth to A_{595} = 0.5, then at 42°C for 2 hours.

To further improve the overproduction of proteins, the region between the RBS and the ATG start codon was investigated to determine the number of nucleotides required for optimal expression. Overproduction of DnaA and Tus (termination utilisation substance) was assessed using a series of vectors constructed by Dr N.E. Dixon (Love *et al.*, in preparation) in which the nucleotide spacing varied from 5 to 11 and 16 bp. Construction of these vectors was achieved by manipulation of restriction endonuclease sites upstream and at the start of the *dnaA* gene in a pCE30 derivative (pPT144) and the formation of an *Nco*I (5´-CCATGG) site at the ATG start codon. These plasmids were transformed into strain AN1459 and are listed in Table 3.1. The vectors pND211-217 and pND230 (*dnaA*$^-$) resulted from the deletion of an *Nco*I fragment (containing the *dnaA* gene) and subsequent recircularisation of the plasmid (Figure 3.5).

For insertion of the *tus* gene into vectors pND211-217 and 230 an *Nco*I site was generated at the start of the gene. A 1158-bp fragment containing *tus* (except for 20 bp at the start) was isolated from pTH311 (Gottlieb *et al.*, 1992) following restriction



Polylinker regions:

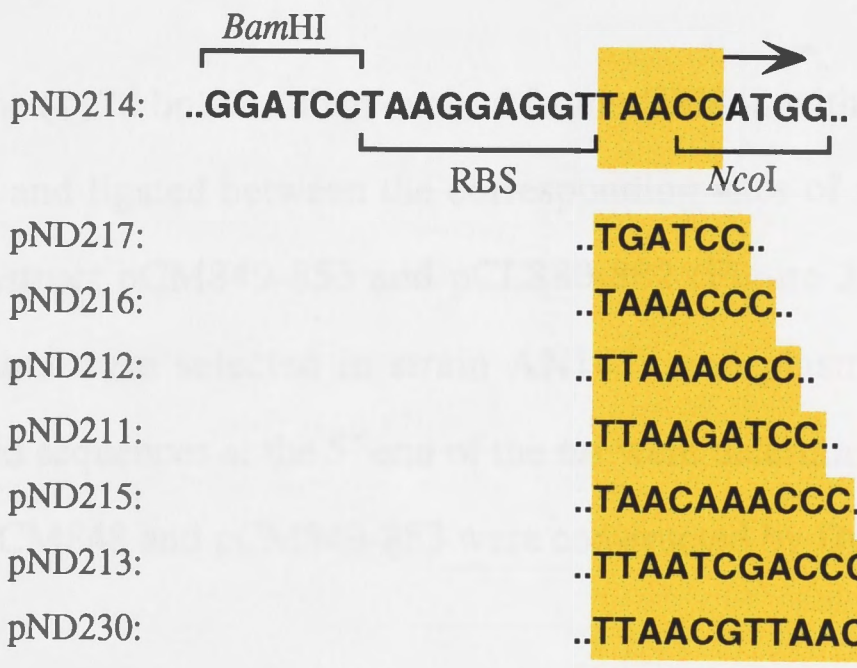


Figure 3.5

The vectors pND211-pND217 and pND230 contain a nine base-pair RBS perfectly complementary to the 3′ end of *E. coli* 16S rRNA varably spaced by means of an AT-rich region from a unique *Nco*I site. The spacer varies in one nucleotide increments from 5 bp in pND214 to 11 bp in pND213 with the exception of pND230 which has a spacing of 16 bp. This region is flanked by *Bam*HI sites for ease of subcloning. Linearization of these vectors with *Nco*I and end-filling with the large fragment of DNA polymerase I results in blunt ends with the sequence CCATG-^{OH}. The terminal ATG can then be employed as a start codon for coding regions of genes inserted in-frame between the blunt ends. This strategy is especially useful in generating N-terminal deletions of protein products (e.g., see Figure 3.6).

endonuclease digestion with *Acc*I and *Sal*I. A ds-DNA oligonucleotide adaptor, a mixture of 5´-CATGGCGCGTTACGATCTCGT-3´ and 5´-CTACGAGATCGT-AACGCGC-3´, was used to add the remainder of *tus* and introduce the *Nco*I site at the start. The vector pMTL20P (Chambers *et al.*, 1988) was digested with *Nco*I and *Sal*I and ligated with the *tus*$^+$ fragment and the oligonucleotide adaptor at 16°C. Following transformation into AN1459, ampicillin-resistant transformants were selected at 30°C and plasmid DNAs were isolated on a small scale. Transformants containing plasmids of the appropriate size (4020 bp) were confirmed to direct overproduction of the desired protein at 42°C, as detected by SDS-PAGE. The plasmid was designated pCM848 (Figure 3.6).

The *tus* gene (1178 bp) was then excised from pCM848 with restriction enzymes *Nco*I and *Eco*RI and ligated between the corresponding sites of vectors pND211-217 and 230 to construct pCM849-853 and pCL880-882 (Figure 3.6). Ampicillin-resistant transformants were selected in strain AN1459 and plasmid DNAs were isolated. Nucleic acid sequences at the 5´ end of the *tus* were determined, and were as expected. Plasmids pCM848 and pCM849-853 were constructed by Dr C.S. Miles.

The strains which overproduce Tus or DnaA (Table 3.1) were grown as described for RSC668 (above). Samples (1 mL) were collected before and after treatment at 42°C and the levels of Tus and DnaA overproduction were assessed by SDS-PAGE. Samples (10 mL) were collected from the strains that overproduced DnaA and the protein was purified to Fraction I as described in Section 3.3.3. The specific activities of DnaA in Fractions I, were determined using the ABC replication assay (Section 3.3.5).

3.3.2 Purification of DnaA Protein

Buffers used in the purification procedure were 25 mM HEPES (pH 7.5), 0.1 mM

Figure 3.6

The scheme for placing an *Nco*I site at the start of the *tus* gene for the cloning of *tus* into the vectors pND211-217 and 230. Part of the *tus* gene (1158 bp), missing the first 20 bp, was excised from pTH311 (Gottlieb *et al.*, 1992) with *Acc*I and *Sal*I. A ds-DNA synthetic oligonucleotide was designed with *Acc*I and *Nco*I overhangs to replace the missing part of *tus* and to introduce an *Nco*I site at the start. These fragments were ligated into pMTL20P (Chambers *et al.*, 1988) that had been digested with *Nco*I and *Sal*I to create the plasmid pCM848 (Figure 3.6). The *tus* gene was removed from pCM848 using restriction endonucleases *Nco*I and *Eco*RI and inserted between the corresponding sites of pND211-217 and 230 (Figure 3.6). The resulting plasmids were pCL880-882 and pCM849-853.

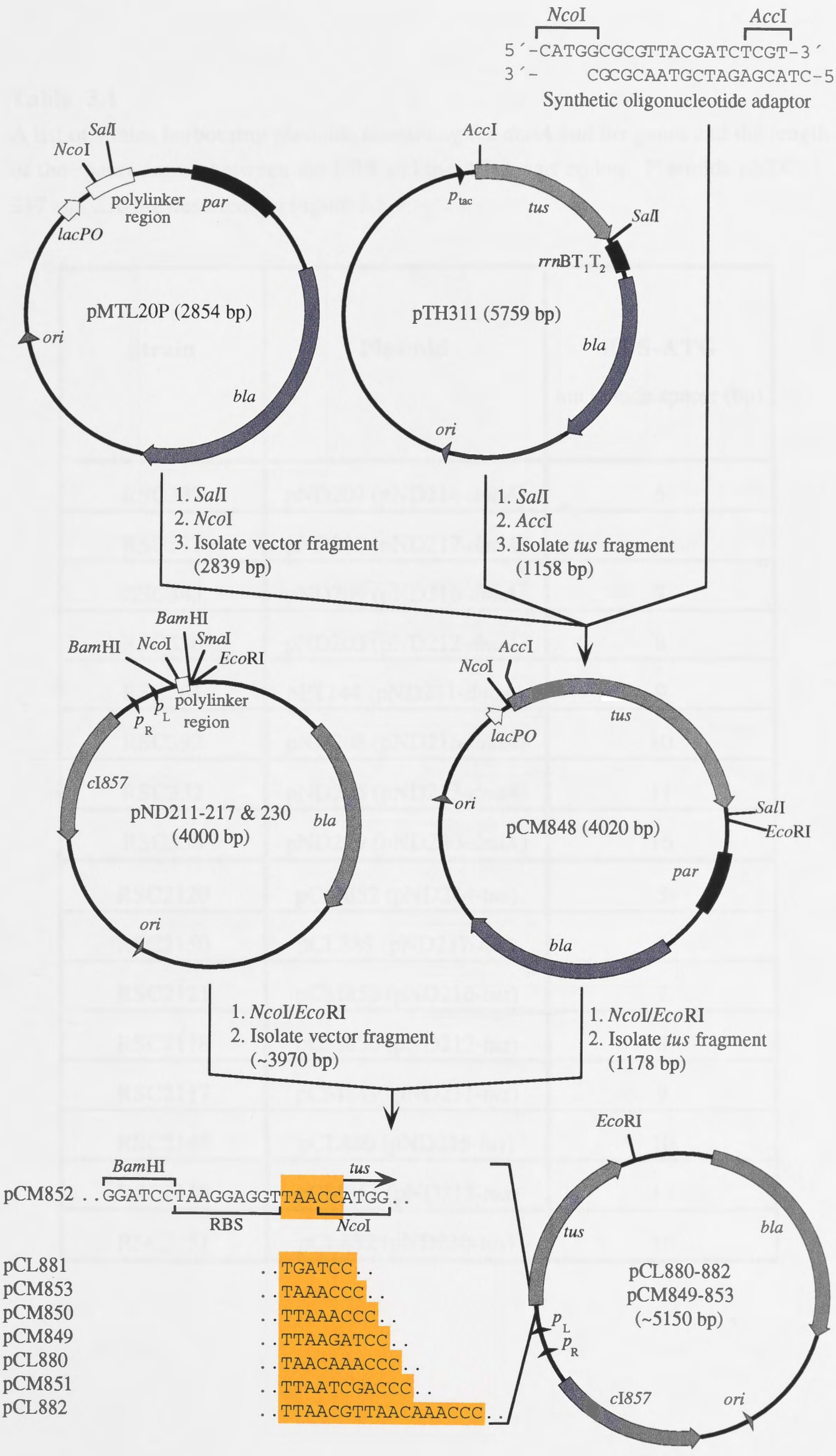

NcoI
AccI
5′-CATGGCGCGTTACGATCTCGT-3′
3′-CGCGCAATGCTAGAGCATC-5′
Synthetic oligonucleotide adaptor
SalI
NcoI
polylinker region
par
lacPO
pMTL20P (2854 bp)
ori
bla
AccI
Ptac
tus
SalI
rrnBT1T2
pTH311 (5759 bp)
bla
ori
1. SalI
2. NcoI
3. Isolate vector fragment (2839 bp)
1. SalI
2. AccI
3. Isolate tus fragment (1158 bp)
BamHI
BamHI
NcoI
SmaI
EcoRI
polylinker region
pR
pL
cI857
pND211-217 & 230 (4000 bp)
bla
ori
AccI
NcoI
tus
lacPO
ori
SalI
EcoRI
pCM848 (4020 bp)
par
bla
1. NcoI/EcoRI
2. Isolate vector fragment (~3970 bp)
1. NcoI/EcoRI
2. Isolate tus fragment (1178 bp)
BamHI
tus
pCM852 . . GGATCCTAAGGAGGTTAACCATGG . .
RBS
NcoI
pCL881 . . TGATCC . .
pCM853 . . TAAACCC . .
pCM850 . . TTAAACCC . .
pCM849 . . TTAAGATCC . .
pCL880 . . TAACAAACCC . .
pCM851 . . TTAATCGACCC . .
pCL882 . . TTAACGTTAACAAACCC . .
EcoRI
tus
bla
pCL880-882
pCM849-853
(~5150 bp)
pL
pR
cI857
ori

Table 3.1

A list of strains harbouring plasmids containing the *dnaA* and *tus* genes and the length of the spacer region between the RBS and the ATG start codon. Plasmids pND211-217 and 230 are described in Figure 3.5.

Strain	**Plasmid**	**RBS-ATG** nucleotide spacer (bp)
RSC337	pND207 (pND214-*dnaA*)	5
RSC373	pND210 (pND217-*dnaA*)	6
RSC343	pND209 (pND216-*dnaA*)	7
RSC330	pND203 (pND212-*dnaA*)	8
RSC261	pPT144 (pND211-*dnaA*)	9
RSC392	pND208 (pND215-*dnaA*)	10
RSC332	pND205 (pND213-*dnaA*)	11
RSC338	pND229 (pND230-*dnaA*)	16
RSC2120	pCM852 (pND214-*tus*)	5
RSC2150	pCL881 (pND217-*tus*)	6
RSC2121	pCM853 (pND216-*tus*)	7
RSC2118	pCM850 (pND212-*tus*)	8
RSC2117	pCM849 (pND211-*tus*)	9
RSC2149	pCL880 (pND215-*tus*)	10
RSC2119	pCM851 (pND213-*tus*)	11
RSC2151	pCL882 (pND230-*tus*)	16

EDTA, 2 mM DTT, 20% (w/v) sucrose, 20 mM spermidine.3HCl and 1 M KCl (lysis buffer) and 25 mM MOPS (pH 6.8), 0.1 mM EDTA, 5 mM β-mercaptoethanol, 15%(w/v) glycerol, 5 mM $MgCl_2$ and 1 mM ATP (Buffer A), containing KCl as specified. A column of P11 cellulose phosphate (fibrous cation exchanger; Whatman) was prepared as directed by the manufacturer and equilibrated with 20 column volumes of Buffer A containing 0.25 M KCl. Sephadex G-200 (Pharmacia) was prepared according to the manufacturer's instructions and equilibrated with 2 column volumes of Buffer A containing 1 M KCl.

Strain RSC668 (AN1459/pND556) was grown at 30°C in 40 L of LB broth containing 50 mg/L ampicillin in a fermentor (New Brunswick Scientific) to A_{595} = 0.5. Overproduction of DnaA protein was induced by rapid temperature shift to 42°C and cell growth was continued at this temperature for 2 hours. After chilling the culture, cells were harvested (Sharples centrifuge), resuspended in 25 mM HEPES (pH 7.5) containing 20% (w/v) sucrose (3 mL/1 g cell paste), poured into liquid N_2 and stored at -70°C. When required, the cell suspension was thawed and diluted in lysis buffer (15 mL/g of cell paste) and cell lysis accomplished with a combination of lysozyme treatment and freeze/thawing. Lysozyme was added to 0.4 mg/mL and the mixture placed at 0°C for 30 min with occasional stirring. The mixture was frozen in liquid N_2, then thawed at 0°C. Cell debris was sedimented by centrifugation (47 800 x *g*; 1 h) and the supernatant collected (Fraction I). Proteins were precipitated by addition of 1.07 volume of 25 mM HEPES (pH 7.5), 0.1 mM EDTA, 5 mM β-mercaptoethanol, 4 M ammonium sulfate. After being stirred for 1 h, the suspension was centrifuged (27 000 x *g*; 30 min). The pellet was either stored at -70°C or dissolved immediately in Buffer A containing 1 M KCl. RNaseA was added (to 50 μg/mL) and the suspension dialysed for 12 h against the same buffer with 2 changes (1 L/6 h). The suspension was clarified by centrifugation (23 500 x *g*; 20 min). The supernatant (Fraction II) was diluted 4-fold with Buffer A, and applied to a column (16 x 2.5 cm) of cellulose phosphate pre-equilibrated in Buffer A + 0.25 M KCl. The column was washed with

Buffer A containing 0.25 M KCl (200 mL); bound protein was eluted with a linear gradient (500 mL) of 0.25 to 1.5 M KCl in Buffer A at a flow rate of 60 mL/h. Fractions containing DnaA were pooled and 1.34 volumes of 25 mM MOPS (pH 6.8), 0.1 mM EDTA, 5 mM β-mercaptoethanol, 4 M ammonium sulfate was added. After being stirred for 1 h, the suspension was centrifuged (27 000 x *g*; 30 min) and the pellet was dissolved in and dialysed against Buffer A + 1M KCl (2 x 1L changes over 12 h). The suspension was clarified by centrifugation (23 500 x *g*; 20 min) and the supernatant (Fraction III) was loaded onto and gel filtered through a column (112 x 2.5 cm) of Sephadex G-200 in Buffer A containing 1 M KCl at a flow-rate of 60 mL/h. Fractions containing DnaA protein, determined by SDS-PAGE, were pooled (Fraction IV).

Purification of DnaA and DnaK from strain RSC625 to the Fraction II stage was performed as described above. The only changes to the procedure were that the cells were grown in 1 L cutures of LBT containing 50 mg/L ampicillin and 7 mg/L chloramphenicol in shaking water baths.

3.3.3 Solubility Experiments

DnaA from strains RSC668 (AN1459/pND556) and RSC625 (AN1459/pPL400-p179) was purified to the Fraction II stage (Section 3.3.2) with the following changes. The ammonium sulfate pellets were resuspended in 25 mM MOPS (pH 6.8), 0.1 mM EDTA, 5 mM β-mercaptoethanol, 15% (w/v) glycerol, 1 M KCl with and without the addition of 5 mM $MgCl_2$ and 1 M ATP. KCl concentrations were reduced to either 0.5 or 0.25 M in the DnaA suspensions by dialysis and the samples were clarified by centrifugation (23 500 x g; 20 min). The supernatants were analysed by SDS-PAGE.

3.3.4 Cloning of the Domain IV Region of DnaA Protein

The plasmid pND209 (Figure 3.7) was digested with *Bst*EII and partially digested with *Nco*I to remove a fragment of 1123 bp. The remaining 5164 bp fragment was isolated and the overhanging ends made blunt with the large fragment of DNA polymerase I and dNTPs. The fragment was re-circularised with ligase and transformed into strain AN1459. Ampicillin-resistant transformants were selected at 30°C and plasmid DNAs were isolated on a small scale. Plasmids of the correct size (5164 bp) were digested with endonuclease *Nco*I to confirm the presence of 1121 and 4043 bp fragments. The resulting plasmid was designated pCL788 (Figure 3.7), in strain RSC2100 (AN1459/pCL788). Strain RSC2100 was grown as described for RSC668 (Section 3.3.1).

3.3.5 DNA Replication Assay

The ABC-primosome assay was used to determine the activity of DnaA protein in the process of DNA replication. All the replication proteins required in the assay were prepared in the research group of Dr N. E. Dixon: the β subunit of DNA polymerase III holoenzyme (3 x 10^6 unit/mg) was prepared by Dr J. L. Beck; DNA polymerase III* (5.5 x 10^5 unit/mg) and SSB (8.8 x 10^6 unit/mg) were obtained from Dr N. E. Dixon; the DnaBC complex (7.9 x 10^4) and primase (1.36 x 10^6 unit/mg; Stamford *et al.*, 1992) were purified by Dr N. P. J. Stamford. Single-stranded R6K-γ2 (M13 A-site) ssDNA template (Masai *et al.*,1990) was a gift from Dr H. Masai (then of DNAX Research Institute of Molecular and Cellular Biology).

The assay was carried out essentially as described Masai *et al.* (1990a) with minor modifications in the components. Reaction mixtures prepared at 0°C contained: 20 mM Tris.HCl pH 7.5; 100 μg/mL BSA; 8 mM DTT; 0.01% (w/v) Brij-58; 8 mM magnesium acetate; 125 mM potassium glutamate; 1 mM rATP; rCTP, rUTP, rGTP

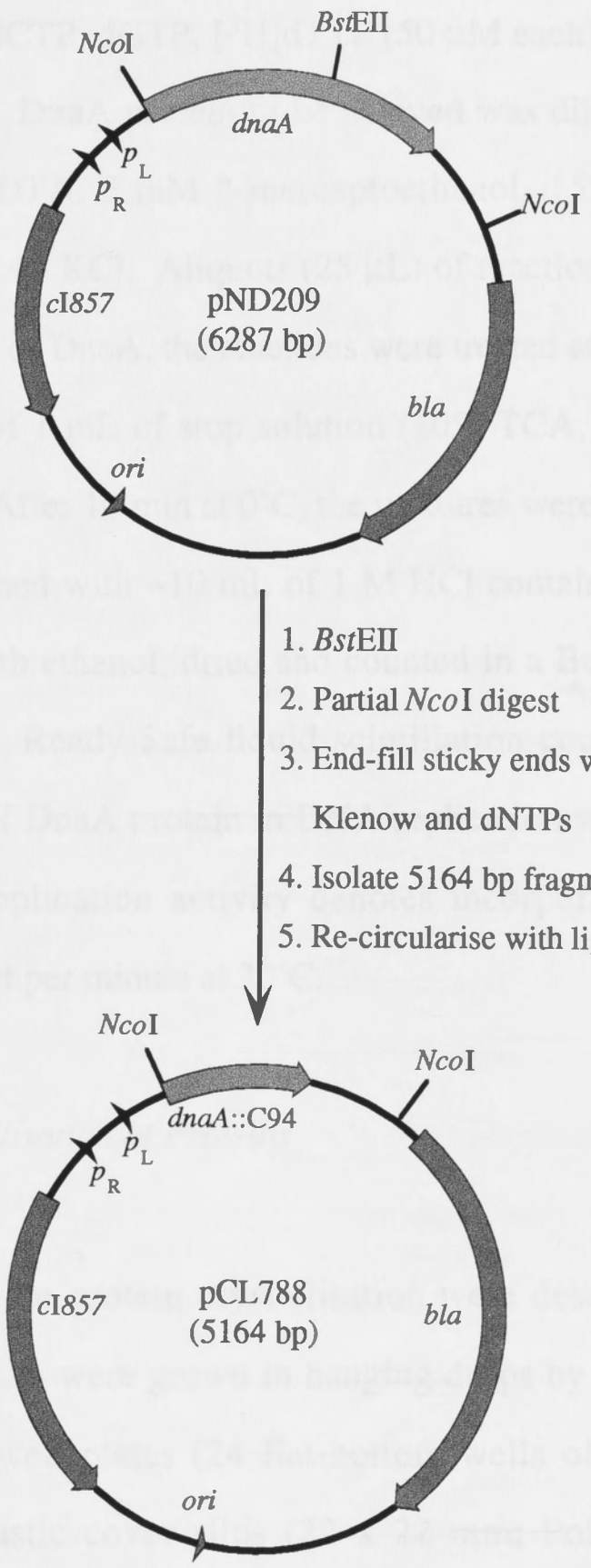


Figure 3.7

A scheme for the construction of a *dnaA* mutant (*dnaA*::C94) which encodes for the C-terminal 94 amino acids of DnaA (Domain IV). The vector pND209 was digested with *Bst*EII and partially with *Nco*I to remove a fragment of 1123 bp. The sticky-ends were made blunt with the large fragment of DNA polymerase I and dNTPs. A fragment of 5164 bp was isolated and re-circularised with ligase to create pCL788.

(250 μM each); 1 μg of SSB; 36 ng of DnaG primase; 390 ng of DnaB.DnaC complex, 100 ng of DNA polymerase III*; 26 ng of β subunit of DNA polymerase III holoenzyme; dATP, dCTP, dGTP, [^{3}H]dTTP (50 μM each) and R6K-γ2 ssDNA (220 pmol, as nucleotide). DnaA protein to be assayed was diluted in 25 mM Na.MOPS (pH 6.8), 0.1 mM EDTA, 5 mM β-mercaptoethanol, 15% (w/v) glycerol, 5 mM $MgCl_2$, 1 mM ATP, 1 M KCl. Aliquots (25 μL) of reaction mixture were prepared at 0°C and after addition of DnaA, the reactions were treated at 30°C for 10 min. Cooling to 0°C and addition of 1 mL of stop solution (10% TCA, 0.1 M $NaPP_i$) terminated replication of DNA. After 10 min at 0°C, the mixtures were filtered through Whatman GF/C filters and washed with ~10 mL of 1 M HCl containing 0.1 M $NaPP_i$. Filters were then washed with ethanol, dried and counted in a Beckman LS 6000 IC liquid scintillation counter. Ready-Safe liquid scintillation cocktail was purchased from Beckman. Activity of DnaA protein in DNA replication was calculated in units (U), where one unit of replication activity denotes incorporation of one picomole of nucleotide into product per minute at 30°C.

3.3.6 Crystallisation of Proteins

The techniques used for protein crystallisation were described by Ollis and White (1990). Protein crystals were grown in hanging drops by vapour diffusion. Limbro tissue culture multi-well plates (24 flat-bottom wells of size 1.7 x 1.6 cm; Flow Laboratory) with plastic cover slips (22 x 22 mm; Polysciences) were used for crystallisation experiments.

Purified DnaA protein for crystallisation was concentrated by addition of 1.34 volumes of 4 M ammonium sulfate in 25 mM MOPS (pH 6.8), 0.1 mM EDTA, 5 mM β-mercaptoethanol. After being stirred for 1 h at 4°C, the suspension was centrifuged (27 000 x *g*; 30 min) and the pellet was dissolved in and dialysed against 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$, 1 mM ATP,

1 M KCl (2 x 1L over 12 h). The suspension was clarified by centrifugation (23 500 x *g*; 20 min). The protein concentration was estimated using the Bradford method (Section 2.14), and the solution was diluted to the appropriate concentration (10-15 mg/mL) for crystallisation. The concentrated DnaA solution was dialysed extensively into 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$, 1 mM ATP containing either 0.25 M KCl or 0.1 M ammonium sulfate.

The "fast" and "medium" screens (Appendix 1) were used to examine a wide range of conditions for crystallisation of proteins. For crystallisation trials, 5 µL of protein solution (5-15 mg/mL) was mixed with 5 µL of the crystallisation buffers (Appendix 1) on ethanol-washed plastic cover slips. The cover slips were then placed over the wells of tissue culture plates containing 1 mL of the same crystallisation buffer and sealed with petroleum jelly. Parallel experiments were set up at 4 and 25°C.

3.3.7 Analysis of Protein Crystals

The appearance, size and morphology of protein crystals were examined using an Olympus SZ PT optical microscope at 4°C. Crystals were photographed using a Pentax camera and Eckachrome 64 ASA film (Kodak). Pictures of crystals presented in this thesis are scans of colour photographs developed by the Instructional Resources Unit, ANU.

The X-ray analysis of crystals was carried out by Dr P. Carr from the Research School of Chemistry using a rotating anode source and an RAXIS-II image plate detector. All X-ray diffraction experiments were performed with crystals sealed in capillary tubes. The crystals were dried, but were in vapour pressure equilibrium with a small amount of the crystallisation solution.

3.3.8 N-Terminal Amino Acid Sequencing of Proteins

N-terminal amino acid sequencing by Edman degradation was performed on a PE Applied Biosystems Procise 494 protein sequencer. The proteins were adsorbed onto ProBlot PVDF membranes and sequenced using standard pulsed liquid blot sequence procedures. The N-terminal sequencing was carried out by the Biomolecular Resource Facility, John Curtin School of Medical Research, ANU.

3.4 Results and Discussion

3.4.1 Overproduction of DnaA protein

The level of overproduction of DnaA protein has been increased dramatically (Lilley *et al.*, in preparation) by changing the GTG start codon to ATG, replacing the natural RBS with a synthetic RBS perfectly complementary to the 3´ end of *E. coli* 16S rRNA and inserting the *dnaA* gene into pPL391 under the control of strong λ p_R and p_L promoters. The strain RSC668 containing the plasmid pND556 (Figure 3.3) directs high-level overproduction of the *dnaA* gene product from the λ promoters upon temperature shift from 30 to 42˚C (Figure 3.8). Based on the activity of DnaA in strain RSC668, 84 000 molecules of DnaA were produced per cell. This is a 3-5 fold increase in DnaA overproduction compared to strain N4830 (pBF1509), based on re-calculation of the number of DnaA molecules per cell (Sekimizu *et al.*, 1988). Although this is a significant amount of protein, other proteins have been overproduced to higher levels in similar vectors (Elvin *et al.*, 1990; Stamford *et al.*, 1992; Love *et al.*, 1996).

Overproduction of DnaA in RSC668 (AN1459/pND556) did not increase beyond 1.5-2 hours of treatment at 42˚C and this was linked with a reduction in the cell density (as measured by A_{595}), which suggests that high levels of DnaA may be lethal to the cell.

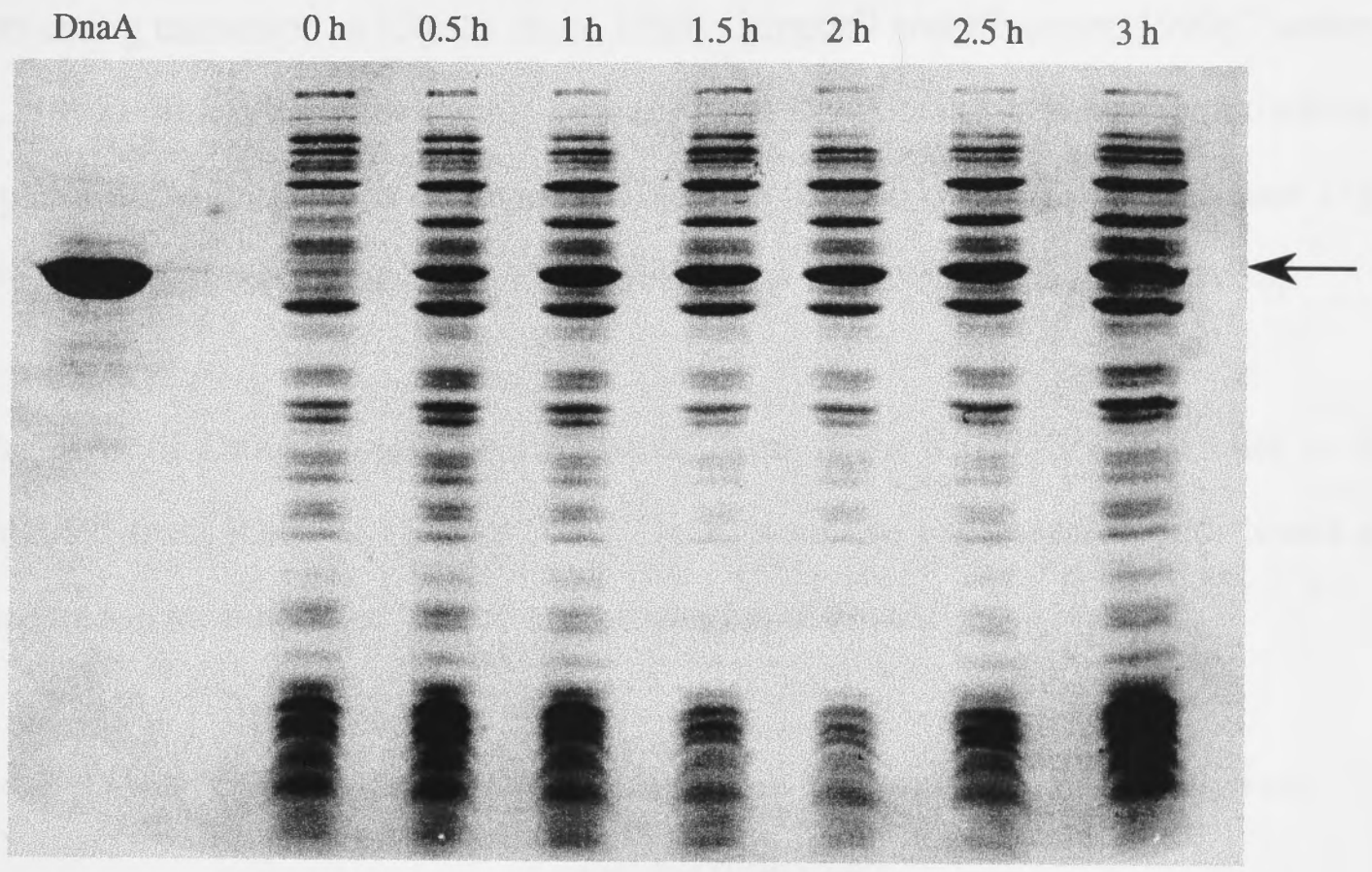


Figure 3.8

Overproduction of DnaA protein directed by strain RSC668 (AN1459/pND556), detected by SDS-PAGE of lysed whole cells. Strain AN1459 containing pND556 (*dnaA*$^+$) was grown at 30°C in LB broth containing ampicillin to A_{595} = 0.5, then shifted to 42°C for 3 hours. Cells in 1 mL samples were harvested before (lane 0 h) and during treatment (lanes 0.5, 1, 1.5, 2, 2.5 and 3 h). They were resuspended in an SDS loading buffer to an A_{595} of 10 and treated at 100°C (2 min). Samples (20 µL) were loaded onto lanes of a 12% SDS-PAGE slab gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue. The arrow indicates the position of DnaA.

In fact, this is easily demonstrated with the strains described here, all of which fail to grow on agar plates at 42˚C. This is not surprising since DnaA autoregulates its own synthesis (Hansen and Rasmussen, 1977; Braun *et al.*, 1985; Atlung *et al.*, 1985; Kucherer *et al.*, 1986) and for a period of the cell cycle the *dnaA* gene is sequestered so that neither RNA polymerase nor *dam* methyltransferase can gain access, thus preventing transcription (Ogden *et al.*, 1988; Campbell and Kleckner, 1990; Theisen *et al.*, 1993). It has been shown that high levels of DnaA result in multiple initiations of replication which have a detrimental effect on the cell cycle (Lobner-Olesen *et al.*, 1989). These observations support the notion that excess DnaA is lethal to cells.

The level of DnaA protein overproduced from strain RSC625 was similar to that obtained from RSC668 (Figure 3.9). DnaK was also expressed at high levels and appeared to have no effect on the overproduction of DnaA.

3.4.2 Effect of the Distance Between the RBS and ATG Start Codon on Gene Expression

Cloning of genes can result in shortening or lengthening of the region between the RBS and the start (ATG) of the gene. This can occur when there is a limited number of useful restriction endonuclease sites available upstream of the start of the gene to be cloned or when the natural RBS is removed during cloning and insert into an expression vector. Therefore, it was crucial to determine whether gene expression was affected by changes in the distance between the RBS and ATG start codon.

Analysis of the overproduction of DnaA from plasmids in which the nucleotide spacing between the RBS and ATG start codon varied from 5 to 11 and 16 nucleotides, suggested that a spacing of 7 nucleotides was optimal (Figure 3.10 and 3.11). The expression of DnaA decreased when the spacer was shortened or lengthened with the exception of DnaA expressed from a spacer of 5 nucleotides. The results indicate that

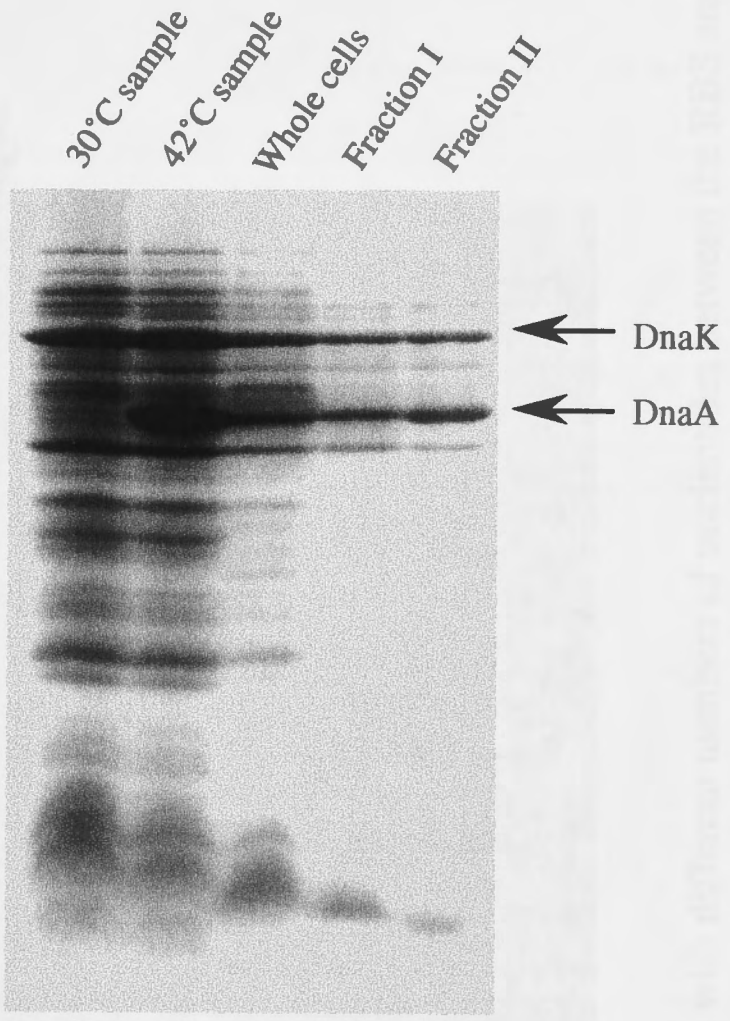


Figure 3.9

Overproduction of DnaA and DnaK and purification to the Fraction II stage. RSC625 (AN1459/pPL400, p179) was grown at 30°C in LB broth containing ampicillin to A_{595} of 0.5 then shifted to 42°C for 2 hours. Cells in 1 mL samples were harvested before (30°C sample) and after treatment (42°C sample), resuspended in an SDS loading buffer to an A_{595} of 10, treated at 100°C (2 min), and samples (20 μL) were loaded onto a 15% SDS-PAGE slab gel. Equivalent volumes of purification samples (whole cells, Fraction I and Fraction II) were treated at 100°C (2 min) prior to loading onto the same gel. Following electrophoresis, the proteins were stained with Coomassie brilliant blue.

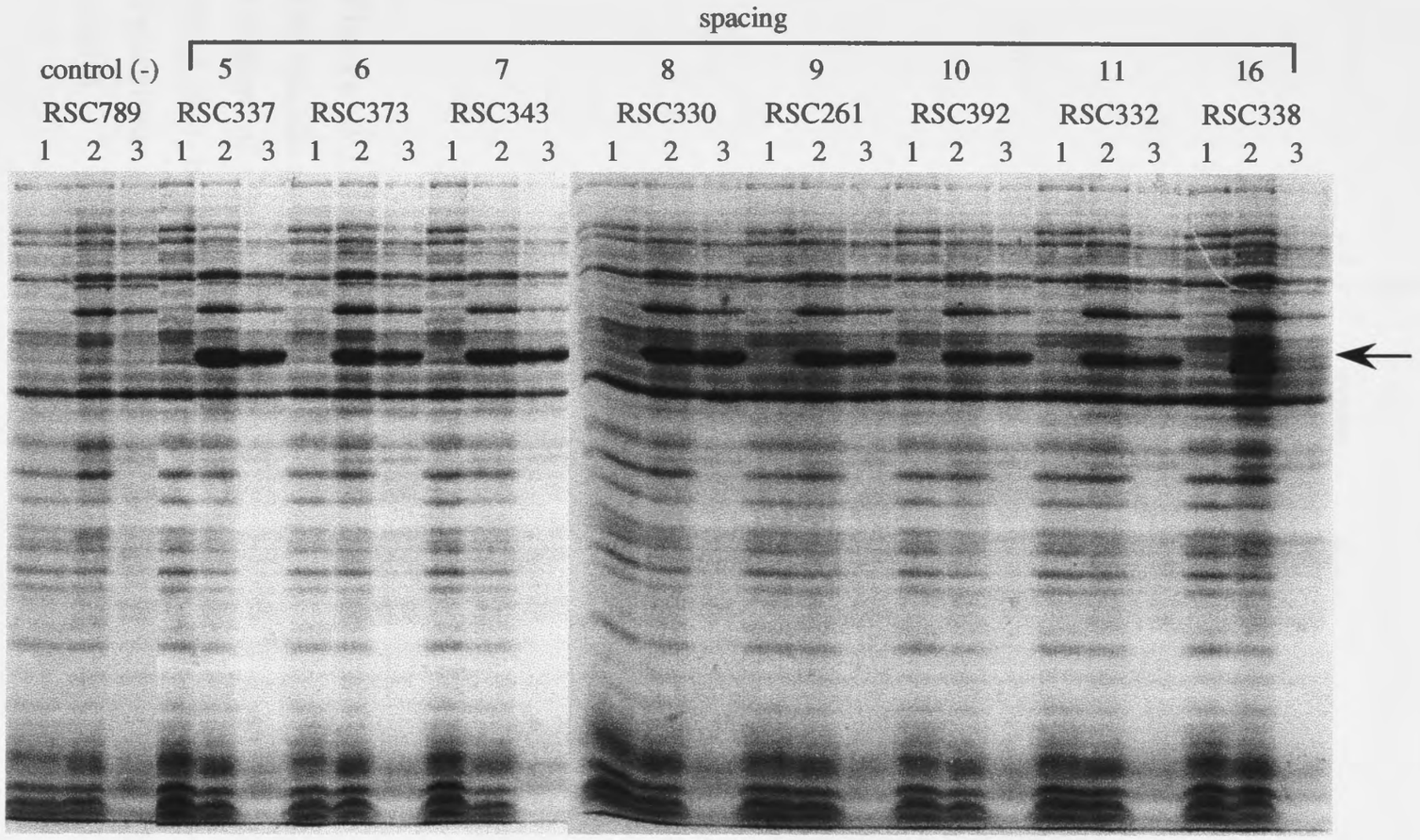


Figure 3.10

Overproduction of DnaA protein directed by strains containing plasmids with different numbers of nucleotides between the RBS and ATG start codon of the *dnaA* gene. Strains (Table 3.1) were grown at 30˚C in LB broth supplemented with ampicillin to A_{595} = 0.5, then shifted to 42˚C for 2 hours. Cells in 1 mL samples were harvested before (lanes 1) and after treatment (lanes 2), and were resuspended in an SDS loading buffer to an A_{595} = 10. The DnaA from 10 mL samples of cells was purified to Fraction I (lanes 3), resuspended in SDS loading buffer and treated at 100˚C for 2 min prior to loading equivalent volumes onto a 12% SDS-PAGE. Following electrophoresis, the proteins were stained with Coomassie brilliant blue. Arrow indicates the position of DnaA protein.

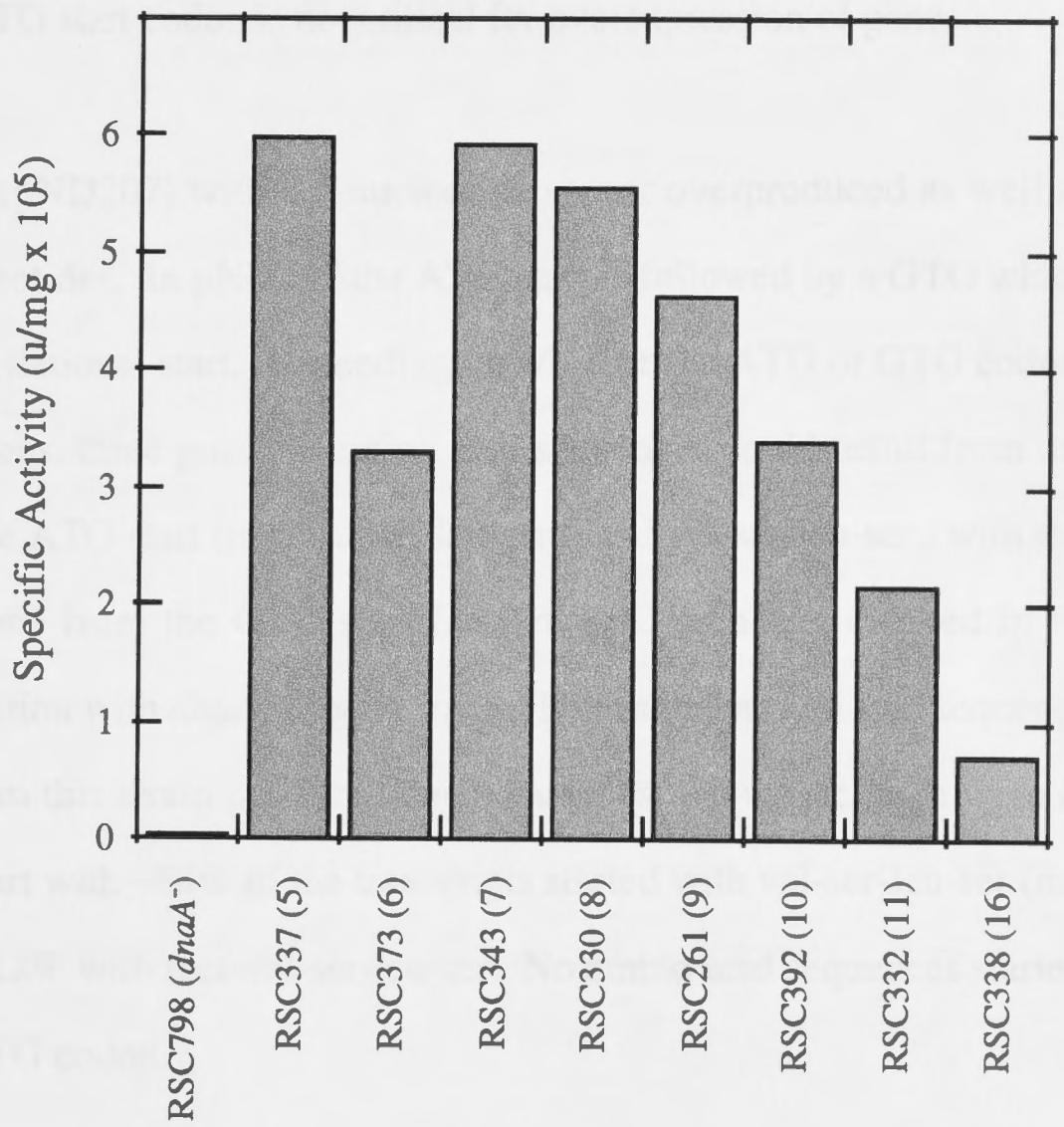


Figure 3.11

The specific activity of DnaA in extracts from strains containing plasmids with different numbers of nucleotides between the RBS and ATG start codon of the *dnaA* gene. DnaA activity was measured in Fraction I extracts (Section 3.2.3) using the ABC replication assay (Section 3.2.4). Strain RSC789 (pPL452 *dnaA*$^-$) was used as a control.

there is only a 2-3 fold difference in the overproduction of DnaA from plasmids with 5-11 nucleotides in the region between the RBS and ATG start. The plasmid containing 16 nucleotides in this region resulted in 8-10-fold less overproduction than a 7-nucleotide spacer. These observations suggest that the nucleotide spacing between the RBS and ATG start codon is not critical for overexpression of genes.

The vector (pND207) with a 5-nucleotide spacer overproduced as well as the vector with 7 nucleotides. In pND207 the ATG start is followed by a GTG which could also act as a translational start. Depending on whether the ATG or GTG codons were used as start codons, three possible amino acid sequences could result from this construct: two from the ATG start (met-val-ser-leu-ser.., and val-ser-leu-ser.. with met cleaved *in vivo*) and one from the GTG start (ser-leu-ser.. with met cleaved *in vivo*; i.e., the normal situation with *dnaA*) (Figure 3.12). N-terminal amino acid sequencing of DnaA purified from this strain produced two sequences. However, both were derived from the ATG start with ~88% of the transcripts started with val-ser-leu-ser (met cleaved *in vivo*) and ~12% with met-val-ser-leu-ser. No amino acid sequences started with serine from the GTG codon.

Similar results were observed for the overproduction of Tus from the same vectors. The production of Tus, as determined by SDS-PAGE, appeared to be very similar in plasmids with a length of 5-11 nucleotides between the RBS and ATG start (Figure 3.13). A length of 16 nucleotides resulted in approximately 2-3-fold less Tus. However, overproduction of Tus is lethal to the cell and as a consequence the cells which overproduce to higher levels stop growing sooner. In the strain that expresses Tus from a plasmid with RBS-ATG length of 16 nucleotides, levels of protein accumulate slower but the cells grow for longer. It was observed that Tus overproduction after 30 minutes resulted in approximately 10-fold less expression when the length between the RBS and ATG was 16 nucleotides.

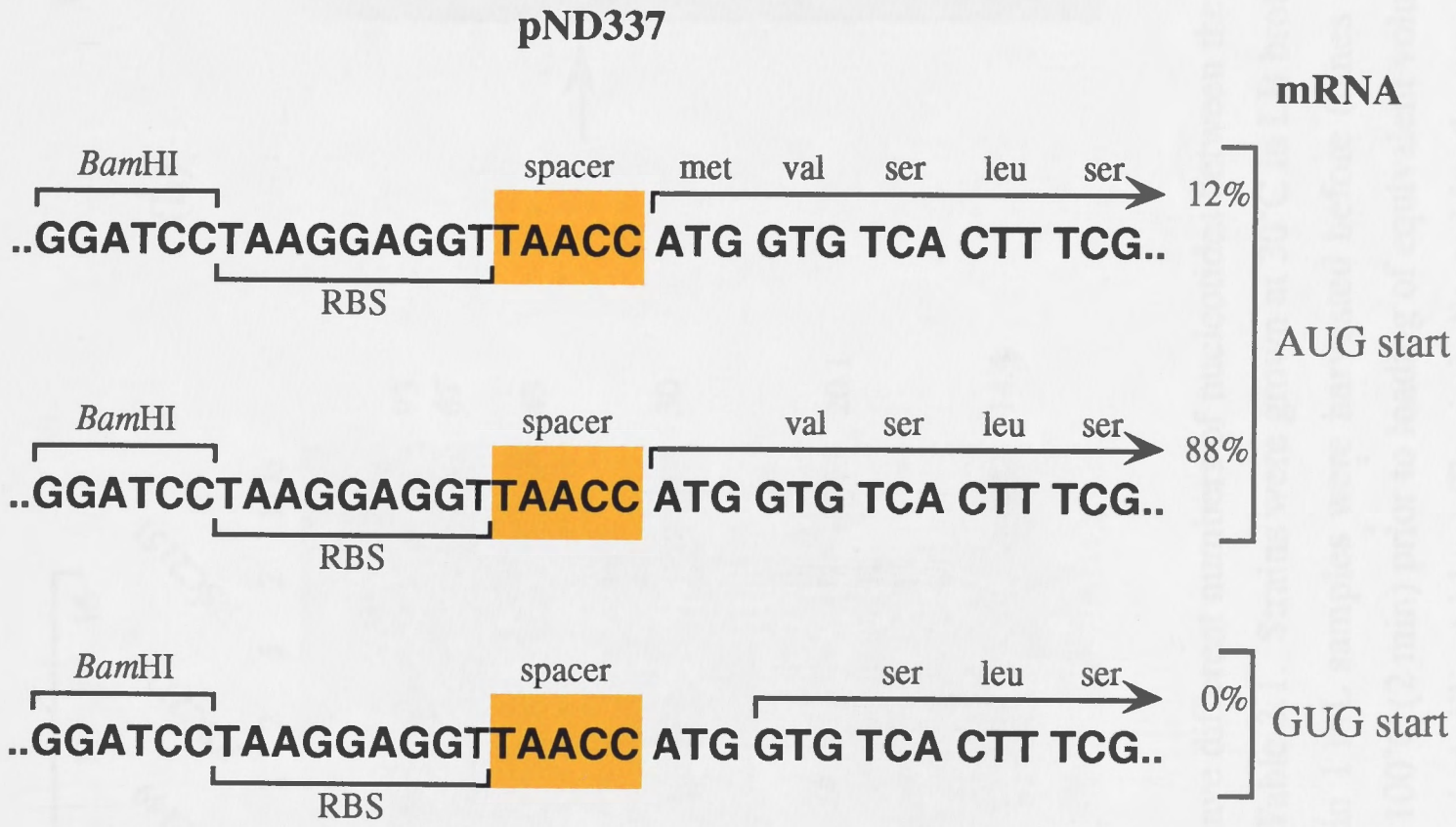


Figure 3.12

The possible N-terminal sequences derived from overproduction of DnaA from pND337 where the ATG start codon if followed by a GTG. N-terminal amino acid sequencing revealed only two sequences, both of which derived from the ATG start codon.

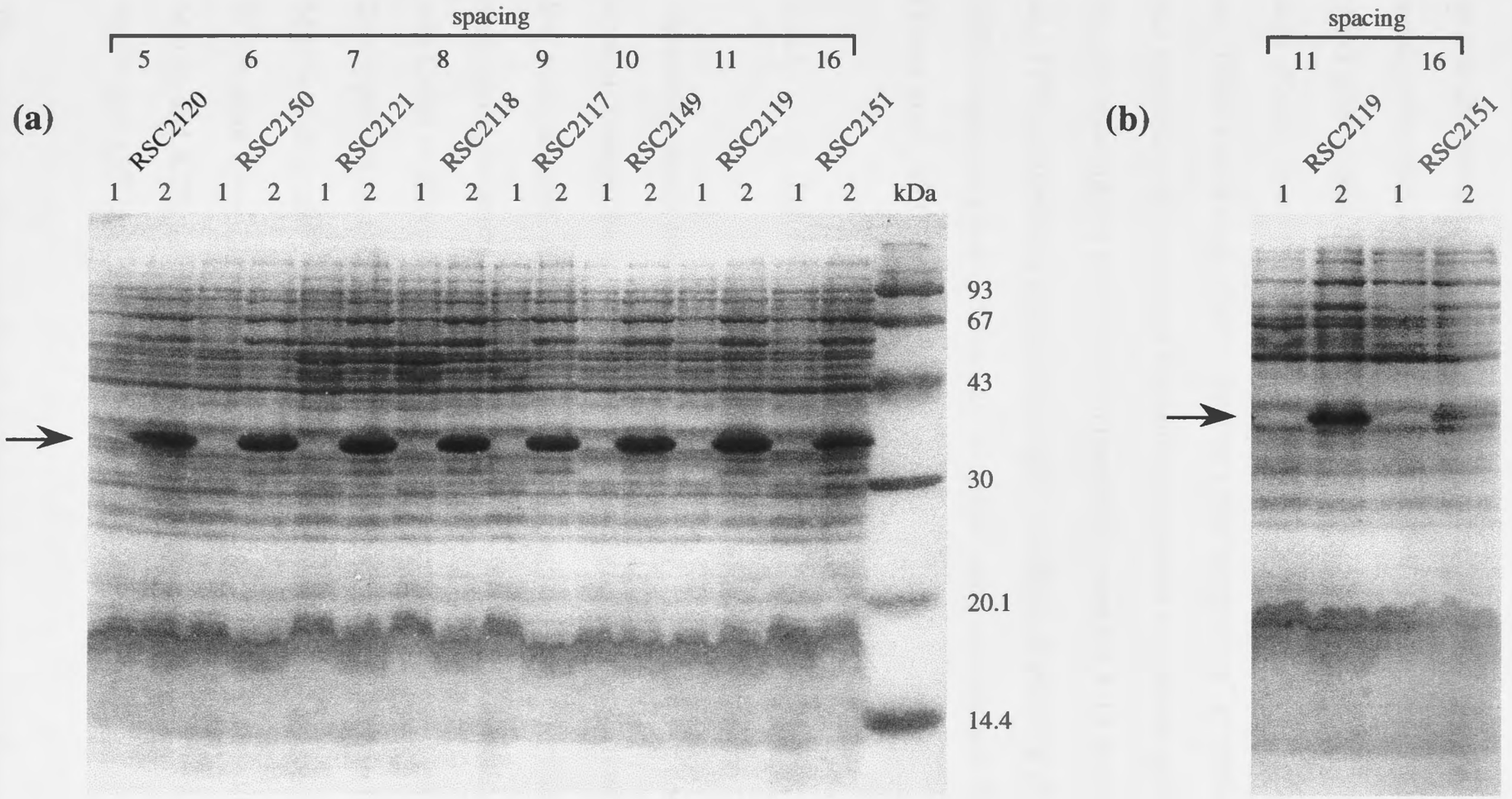


Figure 3.13

The overproduction of Tus from strains containing plasmids which have different numbers of nucleotides between the RBS and the ATG start codon of the *tus* gene. The strains and the plasmids they contain are listed in Table 3.1. Strains were grown at 30˚C in LB broth containing ampicillin to A_{595} = 0.5, then shifted to 42˚C and grown for a further 2 hours. Cells in 1 mL samples were harvested before (lanes 1) and after induction (lanes 2), resuspended in an SDS loading buffer to an A_{595} = 10 and treated at 100˚C (2 min) prior to loading of equivalent volumes onto a 15% SDS-PAGE gel. Following electrophoresis, the proteins were stained with Coomassie brilliant blue. Tus is indicated by the arrow. **(a)** Samples collected were obtained after 2 hours of treatment at 42˚C (lanes 2) **(b)** Samples obtained were collected after 30 min of treatment at 42˚C (lanes 2).

These results reported here are consistent with previously reported studies on the effects of the spacing between the RBS and the ATG start codon. Studies on the *in vivo* expression of *E. coli* fibroblasts and leukocyte interferons (Shepard *et al.*, 1982) and an immunoglobulin heavy chain (Wood *et al.*, 1984) showed that there was only a small difference in translation with spacings of 5-13 nucleotides between the RBS and ATG start codon and a spacing of 9 nucleotides was optimal. However, translation was affected when the spacing was less than 5 or more than 13 nucleotides (Shepard *et al.*, 1982; Wood *et al.*, 1984). Results from "toeprinting" experiments, which measure the formation of translation initiation complexes on various mRNAs, also confirmed that the spacing has little effect on translation over the 5-13 nucleotide range (Hartz *et al.*, 1991). However, there were slight variations depending on the sequence of the RBS, suggesting that ribosomal "slippage" may compensate for less optimal RBSs (Hartz *et al.*, 1991).

3.4.3 Solubilisation of DnaA protein

Successful purification of DnaA protein under non-denaturing conditions required that several obstacles be overcome, notably its aggregation and irreversible precipitation at low ionic strength. Extensive trials to investigate the solubility of DnaA protein at low ionic strength were set up as described (Section 3.3.2). Addition of $MgCl_2$ and ATP was found to alleviate these solubility problems at moderate ionic strength (Table 3.2). DnaA protein was soluble in buffers containing 0.25 M KCl in the presence of 5 mM $MgCl_2$ and 1 mM ATP. DnaA otherwise precipitated irreversibly at ~0.5 M KCl. When pure DnaA protein was precipitated at low ionic strength in the presence of $MgCl_2$ and ATP, the protein could be solubilised by the addition of 0.25 M KCl (with Mg^{2+} and ATP) and remained fully active (Figure 3.14).

The molecular chaperone DnaK is known to associate with DnaA in solution (Malki *et al.*, 1991) and it has been suggested that DnaK maintains DnaA in an active

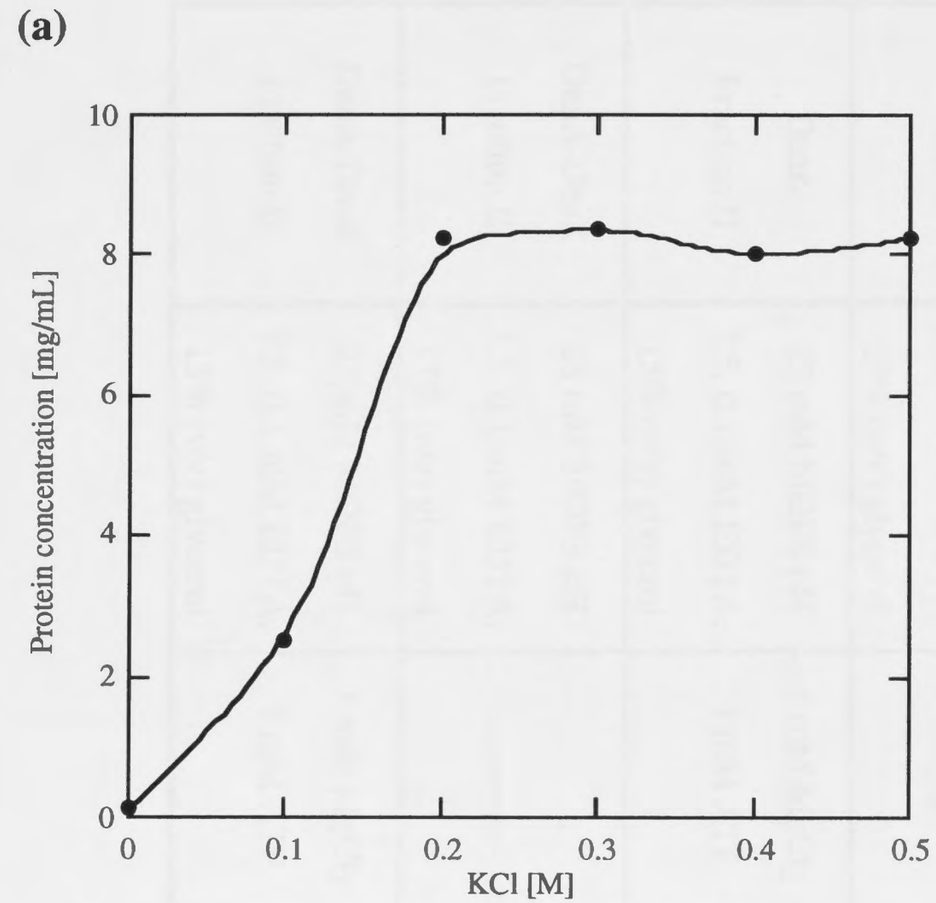


Figure 3.14

(a) Solubility of DnaA protein in the presence of KCl. DnaA (8 mg/mL) protein in 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$ and 1 mM ATP was dialysed against the same buffer. KCl concentrations varied from 0 to 0.5 M by 0.1 M increments. The dialysed solutions were clarified by centrifugation (23 500 x g, 20 min). Protein concentration of supernatants was measured by the Bradford method (Section 2.14).

(b) SDS-PAGE of soluble DnaA protein at different concentrations of KCl. Samples of supernatants (equal volumes) were taken from DnaA solutions that had been dialysed against 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mecaptoethanol, 5 mM $MgCl_2$ and 1 mM ATP at varying KCl concentrations. Following centrifugation, 2 μL samples were loaded onto the gel. The arrow indicates DnaA protein.

Table 3.2

Solubility of DnaA protein in the presence of $MgCl_2$, ATP and DnaK. DnaA was purified to Fraction II from RSC668 by the method described in Section 3.3.2. DnaA co-overproduced with DnaK from strain RSC625 was purified to Fraction II by the same purification method (Section 3.3.2). The KCl concentration was reduced by dialysis and samples were clarified by centrifugation (see Section 3.3.2). Soluble DnaA in the supernatants was analysed by SDS-PAGE. The level of DnaA solubility is represented by the number of (+) symbols, where (++++) indicates that all the DnaA protein was soluble and (+) indicates very little solubility.

Protein Sample	Buffer	Components	[KCl]	Solubility
DnaA Fraction II	25 mM MOPS pH 7.5, 0.1 mM EDTA, 15% (v/v) glycerol	----	0.5 M	++
			0.25 M	+
DnaA Fraction II	25 mM MOPS pH 7.5, 0.1 mM EDTA, 15% (v/v) glycerol	5 mM $MgCl_2$ 1 mM ATP	0.5 M	++++
			0.25 M	++++
DnaA-DnaK Fraction II	25 mM MOPS pH 7.5, 0.1 mM EDTA, 15% (v/v) glycerol	----	0.5 M	++++
			0.25 M	+
DnaA-DnaK Fraction II	25 mM MOPS pH 7.5, 0.1 mM EDTA, 15% (v/v) glycerol	5 mM $MgCl_2$ 1 mM ATP	0.5 M	++++
			0.25 M	++++

conformation for initiation of DNA replication (Skarstad and Boye, 1994). DnaK was found to aid in solubilisation of DnaA in crude fractions at low ionic strengths. When co-expressed with DnaK, DnaA was soluble at 0.5 M KCl (Table 3.2). However, at 0.25 M KCl most of the DnaA had precipitated.

3.4.4 Purification of DnaA protein

Large scale purification of DnaA from the lysate of RSC668 (AN1459/pND556) was carried out by the method described in Section 3.3.3 (Table 3.3). Overproduced DnaA was extracted from the soluble fraction when the cells (10.3 g of cell paste) were lysed by a combination of lysozyme treatment followed by freezing and thawing (Sekimizu *et al.*, 1988). Reproducible separation of DnaA from the bulk of the other proteins was achieved by ammonium sulfate fractionation and ion-exchange chromatography on P11 cellulose phosphate (Figure 3.15). However, some DnaA was lost in the insoluble pellet following ammonium sulfate fractionation (Figure 3.15). DnaA co-eluted with several contaminating proteins from cellulose phosphate chromatography. These impurities were largely removed by gel filtration on Sephadex G-200 (Figure 3.16). DnaA eluted from the Sephadex G-200 column as a single symmetrical peak. Only fractions 56-64 were pooled since fractions 65 onward were contaminated with smaller proteins (see Figure 3.16). This yielded a highly-purified preparation of DnaA (Figures 3.17). A yield of 25% was obtained which could be increased to ~40% by repeating the Sephadex G-200 chromatography step with fractions 65-72 to remove minor contaminants. This compares favourably with a 1% yield obtained by the purification method described by Sekimizu *et al.* (1988). The DnaA isolated by this procedure eluted from a gel filtration (Sephadex G-200) column in a single symmetrical peak at a position expected for a globular 77.6-kDa protein, suggesting that it is monomeric but may have a somewhat asymmetric shape (Figure 3.16).

DnaA and DnaK co-overproduced in strain RSC625 (AN1459/pPL400-p179) were

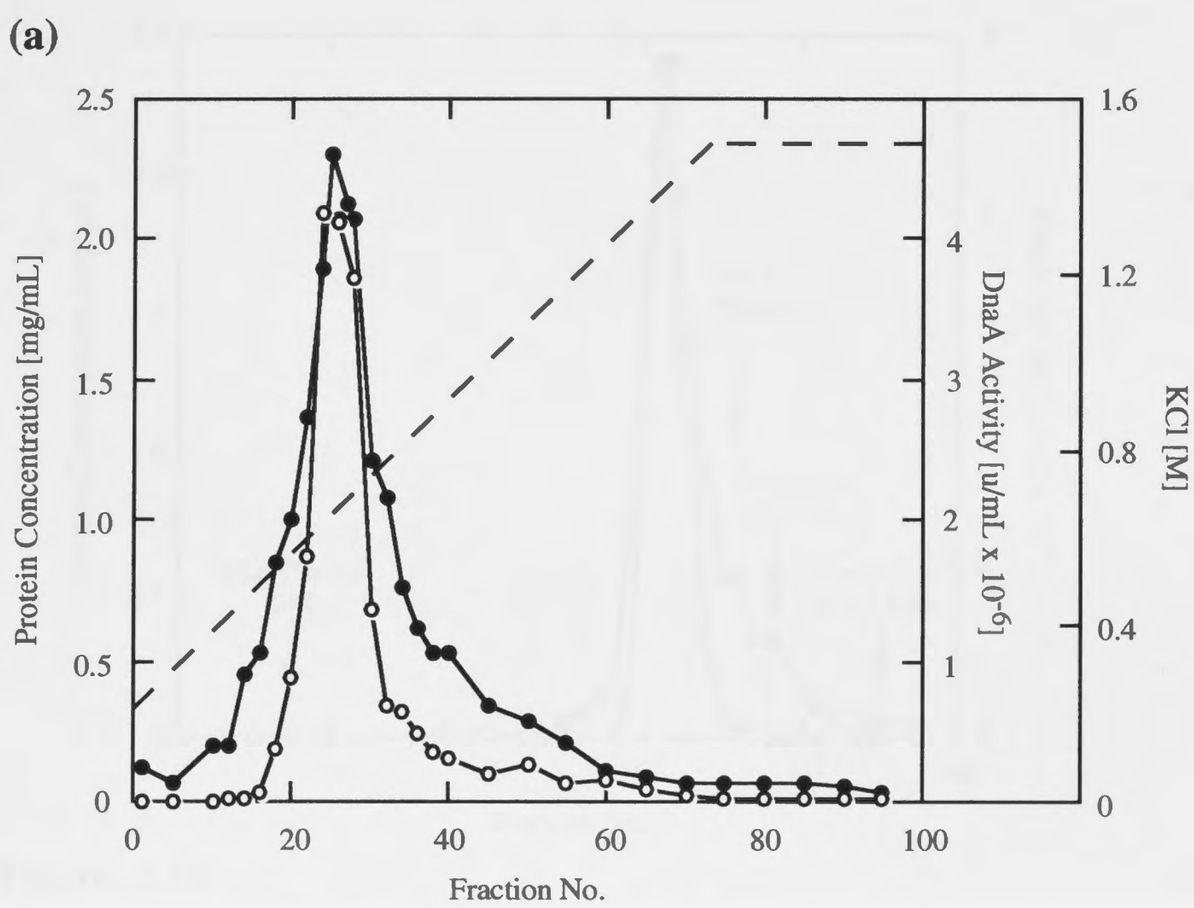


Figure 3.15

(a) The elution profile of DnaA protein from cellulose phosphate chromatography. The details of the chromatography are in Section 3.3.2. Protein concentrations (●) of fractions were determined using the Bradford method (Section 2.14) and the DnaA activity (o) was calculated from results of ABC replication assays (Section 3.3.5).

(b) SDS-PAGE of proteins in whole cells, Fraction I, Fraction II, an insoluble pellet from Fraction II, the cellulose phosphate chromatography flow through and fractions collected from the elution of proteins from the column (numbered as indicated). A volume of 15 μL of the chromatography fractions was loaded onto the gel. DnaA protein is indicated by the arrow.

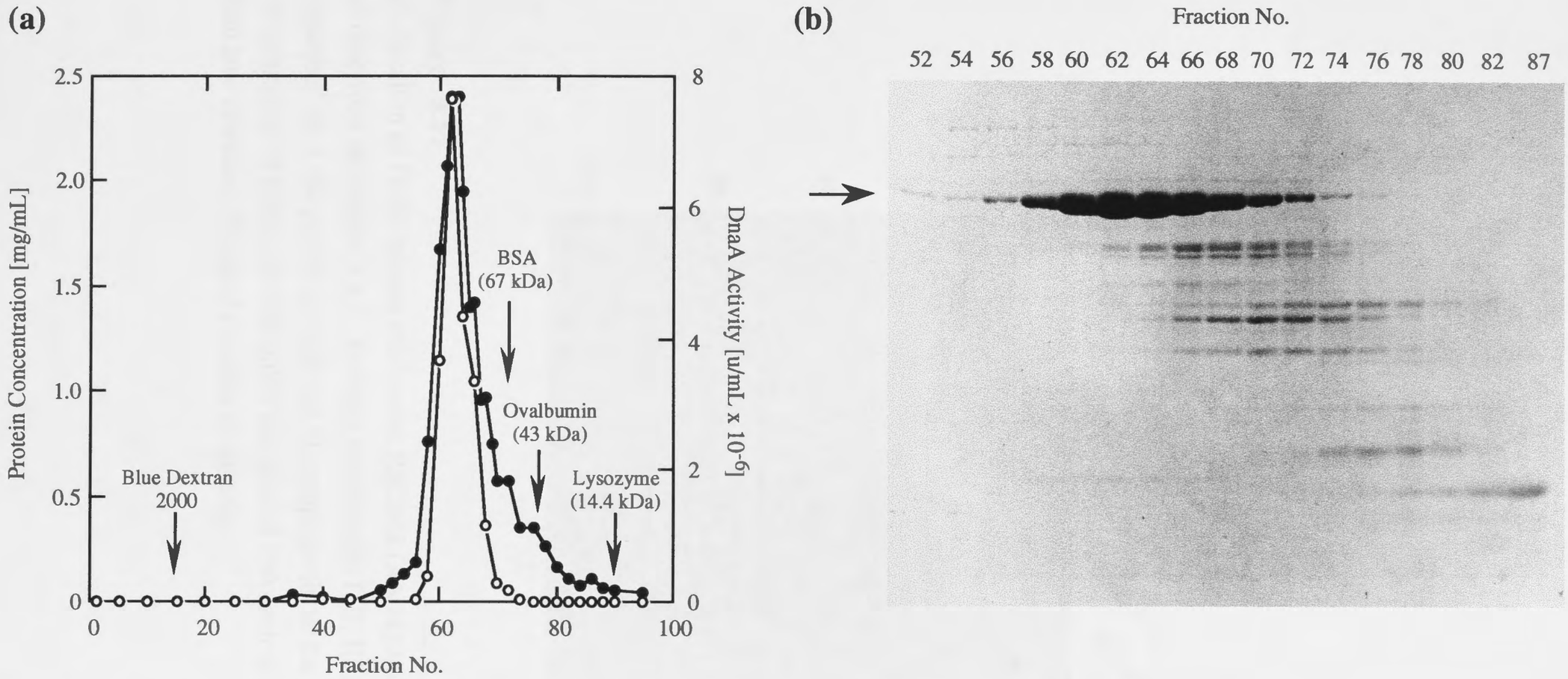


Figure 3.16

(a) The elution profile of proteins separated by Sephadex G-200 gel filtration as described in Section 3.3.2. Determination of protein concentration (●) and DnaA activity (o) is described in the text. Fractions of 6 mL were collected. The elution positions of Blue Dextran 2000, BSA, ovalbumin and lysozyme from a calibration of the Sephadex G-200 column are also shown.

(b) SDS-PAGE of proteins in fractions collected from Sephadex G-200 chromatography. Fractions are numbered as indicated and 15 μL of each fraction was loaded on the gel. The arrow indicates DnaA.

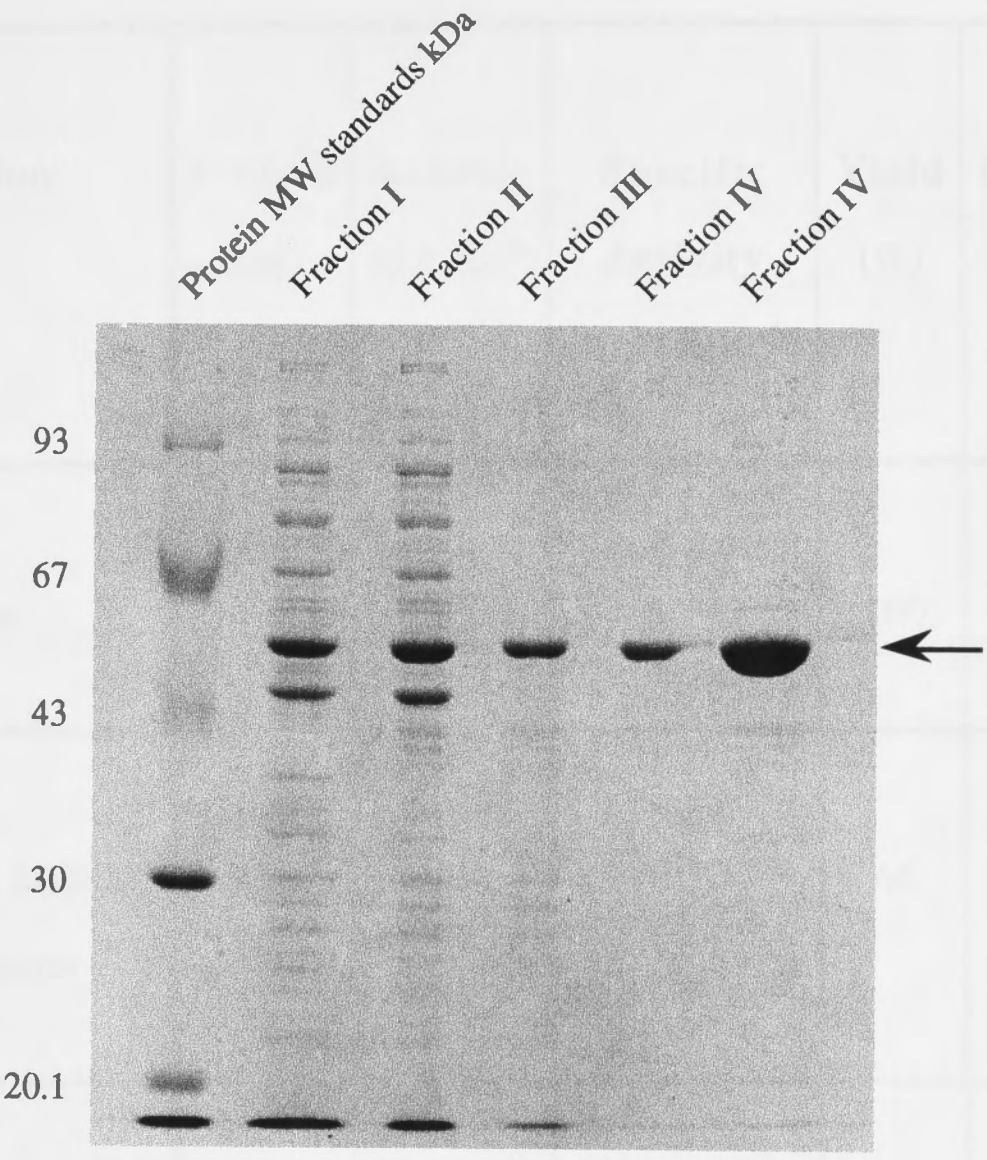


Figure 3.17

Purification of DnaA protein from strain RSC668 (AN1459/pND556) was carried out as described in Section 3.3.2. Proteins in Fractions I, II, III and IV (Table 3.3) were separated by 12% polyacryamide gel electrophoresis in the presence of SDS. The same amount of DnaA (20 000 units) was loaded into each of the first four lanes. The fifth lane contained 30 mg of Fraction IV protein.

Table 3.3

Purification of DnaA protein from 10.3 g of cell paste. DnaA was purified by the method described in Section 3.3.3. Protein concentration was measured by Bradford assays (Section 2.14) and DnaA activity was measured using the DNA replication assay (Section 3.3.5).

Fraction	**Protein** (mg)	**Activity** (u x 10^{-8})	**Specific Activity** (u/mg x 10^{-6})	**Yield** (%)	**Purification** (-fold)
I *Lysate*	1621	8.7	0.54	(100)	(1.0)
II *Ammonium Sulfate Precipitation*	825	5.74	0.70	66	1.3
III *Cellulose Phosphate Chromatography*	132	3.6	2.73	41	5.1
IV *Sephadex G-200 Chromatography*	53.4	2.18	4.08	25	7.6

purified to Fraction II as described in Section 3.3.3 (Figure 3.9). The DnaK protein was in the soluble fraction following lysis by lysozyme and freeze/thaw treatment. The protein precipitated at the same concentration of ammonium sulfate used for precipitation of DnaA and was soluble in Fraction II. The protein solution was then used to investigate the solubility of DnaA at low ionic strength (Section 3.3.2).

The DnaA protein had a specific activity of 4.08 x 10^6 U/mg in the ABC replication ss→RF assay (Section 3.2.5) (Masai *et al.*, 1990). The phage (M13 A-site) DNA used for replication in this assay was single-stranded and contained a single DnaA box in a hairpin loop. In the presence of DnaB and DnaC, replication is dependent on DnaA. From the specific activity, it was calculated that less than two molecules of DnaA were required for complete conversion of single-stranded M13 A-site DNA to a double-stranded product (Figure 3.18). This would suggest that as few as 6-8 molecules would be sufficient for initiation at the four DnaA boxes in *oriC*. However, based on the literature, 20-30 molecules of DnaA are required for *oriC* initiation. Replication assays using plasmids dependent on *oriC* initiation were shown to require 15 molecules to replicate one circle (Fuller and Kornberg, 1983) while electron microscopy of initiation complexes suggest that 20-30 molecules of DnaA are required to form a complex of the size visualised (Funnell *et al.*, 1987; Crooke *et al.*, 1993). On the other hand, Woelker and Messer (1993) have proposed that only 10 molecules of DnaA are required for *oriC* initiation. The number of molecules of DnaA required for the initiation of M13 A-site (a single DnaA box) or *oriC* (four DnaA boxes) suggest that both the process of initiation and the DnaA-DNA complex formed differ. Thus, initiation from a single DnaA box cannot be compared directly to *oriC* initiation. Also, in initiation at *oriC* the DNA is wound around a complex of DnaA molecules and the positions of the DnaA boxes are critical for complex formation. This is probably not the case in initiation on single-stranded M13A-site DNA. In spite of this, however, the mechanism of DnaA-dependent loading of DnaB from the DnaBC complex could be similar.

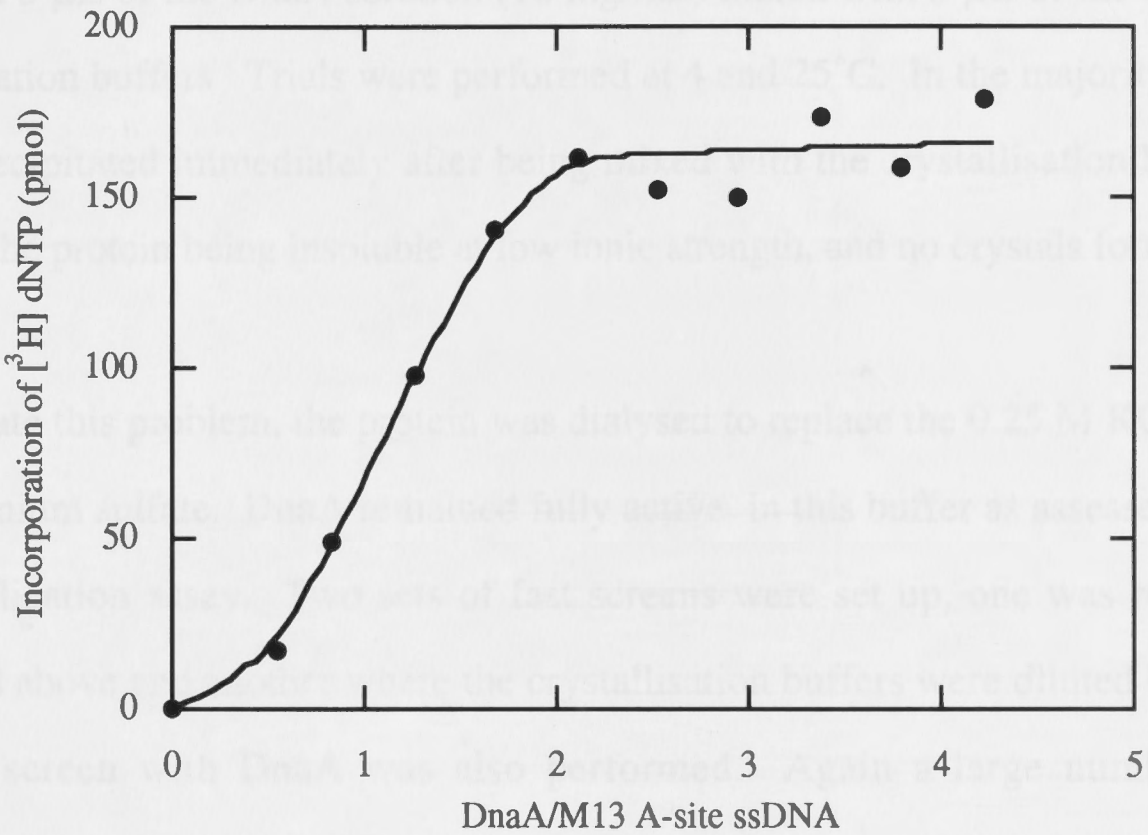


Figure 3.18

The activity of DnaA protein in the ABC replication assay (Section 3.2.5). DnaA protein was titrated into the ABC replication assay and the molecules of DnaA added per circle of M13 A-site ssDNA plotted against the pmol incorporation of [^{3}H] dNTPs. Complete replication of all template circles would produce 220 pmol of product.

3.4.5 Crystallisation of DnaA Protein

The hanging drop method (Section 3.3.6) was used for crystallisation of DnaA. Fast and medium screens (Appendix 1) were used for crystallisation and initial trials were set up using DnaA protein in 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$, 1 mM ATP and 0.25 M KCl. Hanging drops contained 5 μL of the DnaA solution (10 mg/mL) mixed with 5 μL of the fast screen crystallisation buffers. Trials were performed at 4 and 25°C. In the majority of cases, DnaA precipitated immediately after being mixed with the crystallisation buffer as a result of the protein being insoluble at low ionic strength, and no crystals formed.

To alleviate this problem, the protein was dialysed to replace the 0.25 M KCl with 0.1 M ammonium sulfate. DnaA remained fully active in this buffer as assessed with the ABC replication assay. Two sets of fast screens were set up, one was repeated as described above and another where the crystallisation buffers were diluted by half. A medium screen with DnaA was also performed. Again a large number of the crystallisation buffers caused DnaA to precipitate. However, these trials did produce some tiny crystals formed from 0.1 M potassium phosphate pH 7 and 2 M ammonium sulfate. Two small crystals were also observed under other conditions. The first, a badly formed rectangular plate (250 x 125 x 100 μm) which crystallised from 0.05 M Tris.HCl pH 8.5 and 1M ammonium sulfate at 4°C, and the second, a hexagonal plate (250 x 150 x 100 μm) crystallised from 25 mM phosphate pH 7 and 0.3 M ammonium sulfate at 25°C (Figure 3.19). Both crystals took more than six months to form and were birefringent. Only the latter of the two crystals was analysed on the RAXIS-II image detector but did not diffract X-rays, probably due to its small size. As this crystal formed over six months it is possible that it may be a product of proteolysis. Electrospray mass spectrometry was used in an effort to determine whether the crystal contained native DnaA or a proteolytic fragment. However, the size of the crystal did not provide enough sample to determine the mass. Therefore, it is still inconclusive

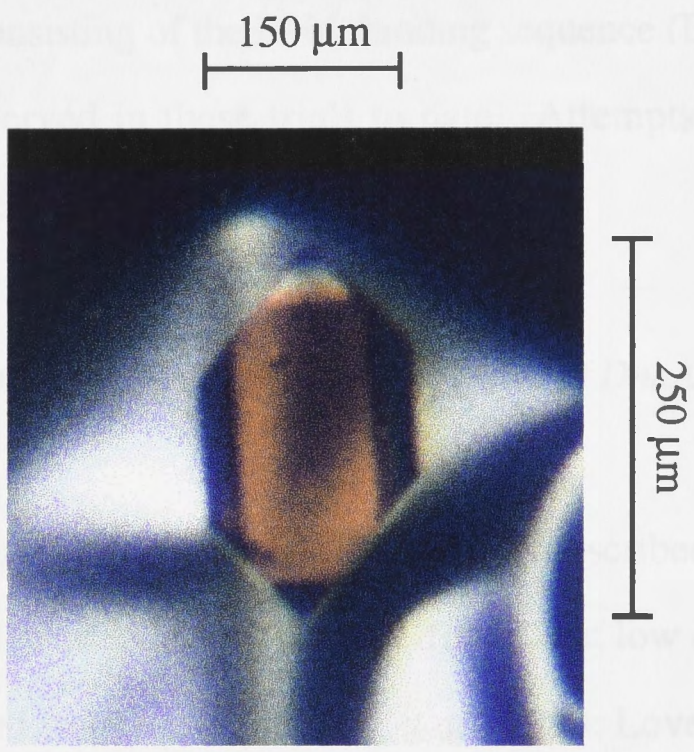


Figure 3.19

A crystal of DnaA protein formed at 25˚C using the hanging drop method. The size of the crystal was 250 x 150 x 100 µm. The DnaA protein (14 mg/mL) was dialysed in 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$, 1 mM ATP and 0.1 M ammonium sulfate. 5 µL of the DnaA solution was mixed with 5µL of crystallisation buffer (25 mM phosphate pH 7, 0.3 M ammonium sulfate) on a cover slip and suspended over the same buffer. The crystal formed over 6 months and did not diffract X-rays.

whether the crystal contained native DnaA.

All crystals formed from solutions containing ammonium sulfate. Thus, extensive investigation of a wide range of ammonium sulfate concentrations (0.6 - 2 M) at pH varying from 4 to 9.5 was performed. Further trials were carried out under these conditions with the addition of ATP, ADP or a dsDNA oligonucleotide (5´-TTATCCACA-3´) consisting of the DNA-binding sequence (DnaA box) for DnaA. No crystal has been observed in these trials to date. Attempts to reproduce the above crystals have also been unsuccessful thus far.

3.4.7 Overproduction and Purification Trials of DnaA::C94

DnaA::C94 was overproduced in strain RSC2100 as described in Section 3.3.1 (Figure 3.20). The level of overproduction was significant, but low compared to many of the proteins overproduced in our group (Elvin *et al.*, 1990; Love *et al.*, 1996). Currently the overproduction of this mutant protein is being improving by sub-cloning its gene into the vector pETMCSI (Figure 2.2).

Overproduction of DnaA::C94 in large scale (1 L) cultures was used to initiate purification trials. The cells were lysed by a combination of lysozyme and freeze/thawing (as described in Section 3.3.3) and following centrifugation (47 000 x *g*, 1 h) DnaA::C94 was located in the soluble fraction (Figure 3.20). Ammonium sulfate fractionation trials are underway which will be followed initially by cellulose phosphate chromatography. This fragment of DnaA is predicted to have a high positive charge at neutral pH, and to contain the DNA-binding motif of the protein.

3.5 Conclusions

The aim of the work presented in this Chapter was to improve the overproduction of

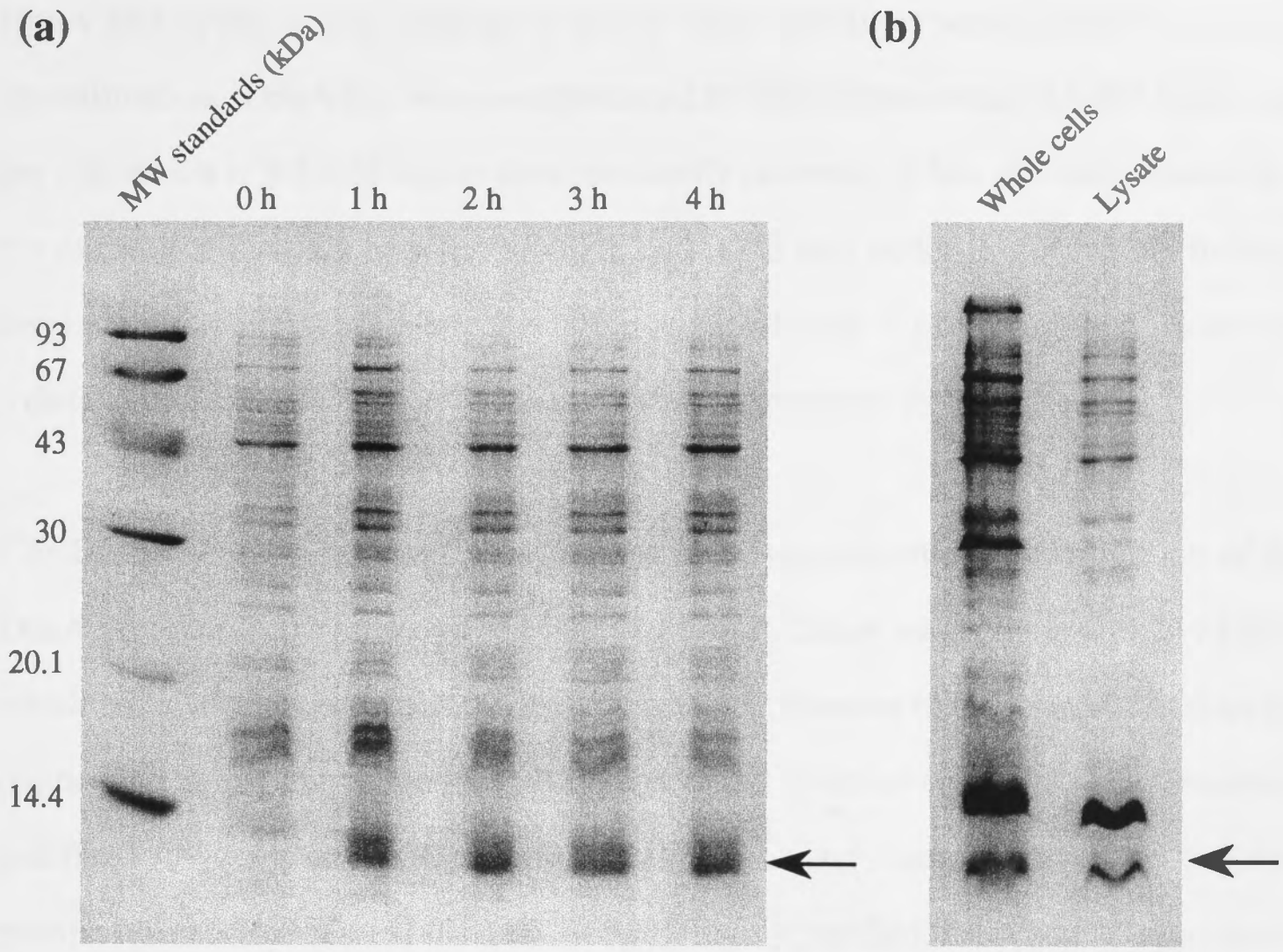


Figure 3.20

(a) Overproduction of DnaA::C94 protein directed by strain RSC2100 (AN1459/pCL788), detected by SDS-PAGE of lysed whole cells. Strain AN1459 containing pCL788 (*dnaA*::C94$^+$) was grown at 30°C in LB broth containing ampicillin to A_{595} = 0.5, then shifted to 42°C for 4 hours. Cells in 1 mL samples were harvested before (lane 0 h) and during treatment (lanes 1, 2, 3 and 4 h), were resuspended in an SDS loading buffer to an A_{595} of 10 and treated at 100°C (2 min). Samples (20 µL) were loaded onto lanes of a 12% SDS-PAGE slab gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue. The arrow indicates the position of DnaA::C94.

(b) An SDS-PAGE gel of overproduced DnaA::C94 before (whole cells) and after (lysate) lysis of whole cells by a combination of freeze/thawing and lysozyme treatment. Strain RSC2100 (AN1459/pCL788) was grown at 30°C in LB broth containing ampicillin to A_{595} = 0.5, then shifted to 42°C for 4 hours. Cells were harvested by centrifugation (10 000 x *g*, 10 min), resuspended in 15 mL/g of lysis buffer (25 mM HEPES pH 7.5, 0.1 mM EDTA, 2 mM DTT, 20%[w/v] sucrose, 20 mM spermidine.3HCl and 1M KCl), then treated with lysozyme (0.4 mg/mL) for 30 min. The mixture was frozen in liquid N_2, then thawed at 0°C. Cell debris was sedimented by centrifugation (47 800 x *g*, 1 h). SDS loading buffer was added to equivalent amounts of samples collected before (Whole cells) and after (Lysate) cell lysis samples were treated at 100°C (2 min) prior to loading onto a 15% SDS-PAGE gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue. The arrow indicates the position of DnaA::C94.

DnaA and devise a new strategy to purify large quantities which could be used in crystallisation. DnaA has been overproduced to levels representing 84 000 molecules per cell which is 3-5 fold higher than previously reported. It has also been shown that the nucleotide distance between the RBS and ATG start codon is not critical for high level expression of genes when this distance varied from 5-11 nucleotides. However, a distance of 16 nucleotides reduced the level of expression by 8- to 10-fold.

Conditions have been identified which prevent aggregation and precipitation of the DnaA protein. In the presence of $MgCl_2$ and ATP, DnaA was soluble at 0.25 M KCl, which has enabled a new strategy for large scale purification to be devised that does not require the use of guanidine.HCl. This procedure involves cell lysis using lysozyme and freeze/thawing, ammonium sulfate fractionation and chromatography on cellulose phosphate and Sephadex G-200, resulting in highly purified, fully active, monomeric DnaA in an overall yield of 25%. It was shown that less than two molecules of DnaA (purified in this manner) was required to replicate one circle of M13 A-site ssDNA.

Attempts to crystallise DnaA have only yielded a couple of crystals, one of which was very badly formed and the other did not diffract X-rays. Both were very small and formed after six months, which may suggest that they were crystals of proteolytic or degradation products. Mass spectrometry was used as an approach to determine if the crystals contained native DnaA. However, the crystals were too small to produce enough sample for detection. Further crystallisation attempts are underway using refined conditions.

The C-terminal 94 amino acids (DnaA::C94) of DnaA have been identified as the DNA binding domain (Roth and Messer, 1995). This region of the gene has been cloned, overproduced and purification is currently underway. The mutant DnaA::C94 was found to be soluble following lysis by lysozyme treatment and freeze/thawing. Once pure, attempts to solve the structure of this domain will be made using either X-ray

crystallography or multinuclear NMR.

CHAPTER 4

STRUCTURAL AND FUNCTIONAL STUDIES OF THE DnaA PROTEIN

4.1 Aims and Significance

The focus of work outlined in this chapter is to confirm that purified DnaA is monomeric and to determine the number of ATP molecules bound to DnaA in solution. Sedimentation equilibrium experiments were used to determine the oligomeric state of the protein in solution. Electrospray ionisation mass spectrometry (ESI-MS) was used to determine the monomer molecular weight of DnaA and proved very useful in providing information on post-translational modification of the protein.

For determination of the number of ATP molecules bound to DnaA, the UV protein spectrum of the protein was simulated using data obtained from the UV spectra of N-acetyl-L-tryptophan ethyl ester, N-acetyl-L-tyrosine ethyl ester and ATP. DnaA spectra simulated with zero, one or two molecules of ATP were compared with the observed spectrum to estimate the number of ATP molecules bound to the protein.

4.2 Introduction

The biochemical function of proteins could be better understood from structural information. There are many techniques available for use in characterising and studying macromolecules such as proteins. These include sedimentation equilibrium, ESI-MS and UV-vis spectroscopy which were all employed to further study of DnaA. Various techniques were used to show that DnaA is monomeric, to confirm the molecular weight, to obtain information about its shape, and to determine the extent of ATP binding.

DnaA protein previously purified was found to exist in two forms, one monomeric (active) and the other aggregated (inactive). Thus, it was necessary to confirm that DnaA purified by the method described in Section 3.3.3 was monomeric in solution. Sedimentation studies are commonly used in the characterisation of proteins. Using the

appropriate techniques, information on the molecular weight, density and shape of a protein can be obtained. During ultracentrifugation, particles are made to move by centrifugal force and the distribution in concentration of the particles in the tube is determined one or more times. There are two types of sedimentation experiments, sedimentation velocity and sedimentation equilibrium. Measurements of molecules while they are moving along the centrifugal axis are known as sedimentation velocity experiments and result in determination of the sedimentation coefficient, a quantity that may provide information about the molecular weight and the shape of the molecule. Under conditions where the distribution of the molecule no longer changes, the molecules are at sedimentation equilibrium (sedimentation rate is balanced by diffusion), and this type of measurement yields data concerning molecular weights and composition.

Mass spectrometry is becoming one of the choice methods for characterisation of large, fragile biomolecules. In use of older techniques, macromolecules such as proteins would decompose before they could be vaporised and even if vaporisation were possible the molecules would be destroyed by ionisation. New technology has overcome such problems and two gentle ionisation methods are now widely available: matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) (Karas and Hillenkamp, 1988) and electrospray ionisation mass spectrometry (ESI-MS) (Meng *et al.*, 1988; Fenn *et al.*, 1989). These techniques have revolutionised biological mass spectrometry. Both allow efficient production of gas-phase ions from intact, large, fragile, polar and thermally-labile biopolymers such as proteins, oligonucleotides, oligosaccharides and carbohydrates. In terms of molecular weight determination, mass spectrometry is 2-3 orders of magnitude more accurate than the techniques previously used (e.g., SDS-PAGE, gel filtration and ultracentrifugation). Mass spectrometry has other advantages over more traditional techniques, e.g. being able to sequence blocked proteins and correct errors in protein sequences derived from DNA or cDNA, define N- and C-terminal sequence

heterogeneity and identify post-translational modifications such as deamidation, isoaspartyl formation, phosphorylation, oxidation, disulfide bond formation and glycosylation.

All mass spectra acquired in this thesis were obtained from ESI-MS. The utility of ESI-MS for the determination of extremely accurate (up to 0.001% mass accuracy) molecular masses of large, fragile molecules, such as proteins, has been clearly demonstrated (Fenn *et al.*, 1990). Due to the formation of multiply charged molecular ions during the ionisation process, large proteins can be analysed on quadropole mass spectrometers with limited mass to charge (m/z) ranges at low concentrations (typically 1-20 μM). Since detection of a protein by ESI-MS is independent of the nature of modifying groups, this technique is an ideal tool to aid in the examination of post-translational modifications.

In theory, a solution of analyte molecules (e.g., protein) is passed through a needle that is kept at high electrical potential. At the needle tip, the solution disperses into a mist of small, highly-charged droplets containing protein molecules. The droplets evaporate rapidly and multiply protonated protein molecules are released into the gas phase. The ionisation of liquids into the gas phase is very gentle and probably occurs with at least one solvation shell still surrounding the protein. Once in the gas phase, the protein molecules are leaked into the vacuum where they are separated and detected according to their m/z ratio. A countercurrent flow of N_2 bath gas is employed prior to proteins entering the vacuum chamber to free the residual solvent (drying effect). Multiple charging (on average, one charge per kDa) leads to the characteristic appearance of ESI-MS spectra. This allows mass analysers with a limited mass-to-charge range to analyse quite large molecules.

Using ESI-MS/MS (tandem mass spectrometry), a single species in a mixture can be further analysed and in the case of small peptides, information concerning the primary

amino acid sequence can be obtained. The first mass analyser is used to select the ion of interest. This ion is passed into a collision cell where it collides with a collision gas, typically argon. The process of fragmentation which then occurs is known as collision activated dissociation (CAD). The second mass analyser separates and measures the masses of the daughter ions produced. Fragmentation occurs mainly at the amide bonds of the peptide backbone as shown in Figure 4.1. Tandem mass spectrometry is primarily used to obtain structural information about the molecule of interest.

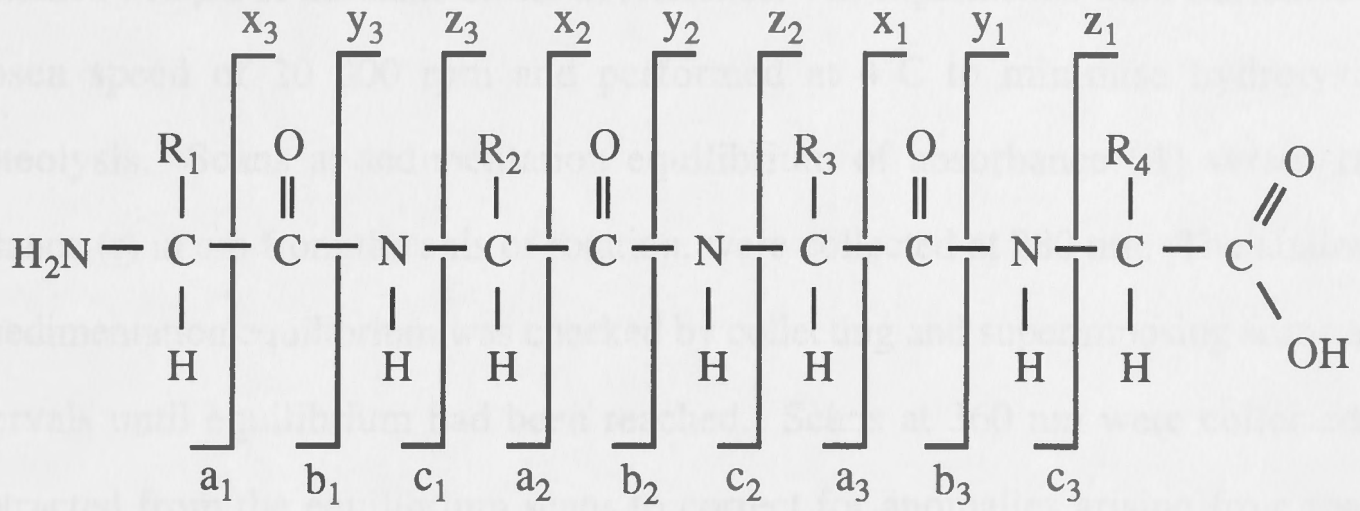
The *E. coli* DnaA replicator initiator protein is known to bind ATP at very high affinity (K_d = 0.03 μM), and this ATP is slowly hydroysed to ADP. Only the ATP-bound form is active (Sekimizu *et al.*, 1987) and the ATP binding site has been located (Figure 3.2) (Koonin, 1993). A second low-affinity binding site for ATP has also been proposed by Sekimizu *et al.* (1987). The DnaA purification methods described here were carried out in the presence of ATP as it was found that ATP and $MgCl_2$ prevented protein aggregation and low ionic strength precipitation. Thus, a determination of the number of ATP molecules bound to DnaA purified by these methods was required for comparison with known data. It was considered possible to determine the number of ATP molecules bound to DnaA by comparison of its UV spectrum with predicted spectra of DnaA with zero, one and two molecules of ATP (based on the UV spectra of ATP, N-acetyl-L-tryptophan ethyl ester and N-acetyl-L-tyrosine ethyl ester).

4.3 Materials and Methods

4.3.1 Sedimentation Equilibrium of Proteins

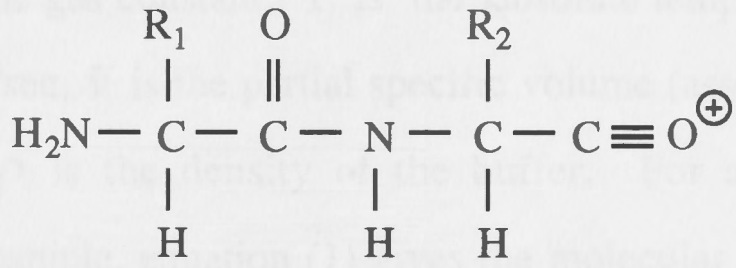
Sedimentation equilibrium experiments were performed in a Beckman Optima XL-A analytical ultracentrifuge using 12-mm path-length cells with carbon-filled double sector centrepieces. The solution sector contained 100 μL of sample of appropriate

(a)

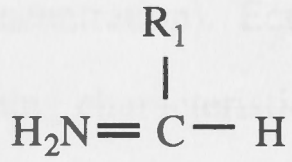


(b)

b_2



immonium ion



y''_2

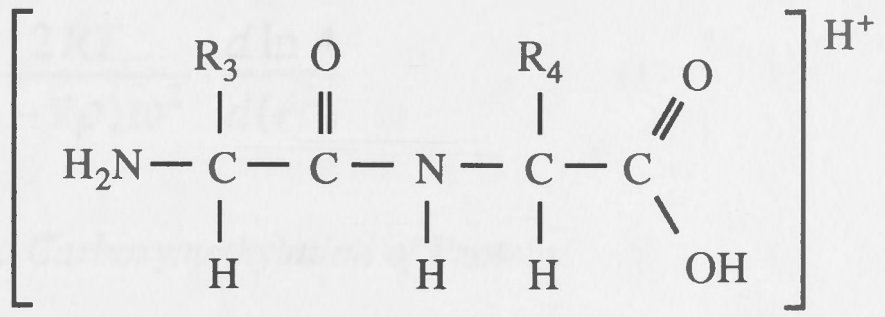


Figure 4.1

(a) An illustration of the characteristic decompositions that linear protonated peptides undergo in CAD processes. The nomenclature used in this figure is that proposed by Roepstorff and Fohlman (1984).

(b) Selected protonated ions observed in tandem mass spectra of peptides.

concentration (between 0.5 and 1 mg/mL) dissolved in buffer; the solvent sector contained 110 μL of the same buffer as reference. All experiments were carried out at a chosen speed of 20 000 rpm and performed at 4°C to minimise hydrolysis or proteolysis. Scans at sedimentation equilibrium of absorbance (*A*) versus radial distance (r) in cm from the axis of rotation, were collected at 280 nm. The attainment of sedimentation equilibrium was checked by collecting and superimposing scans at 2 h intervals until equilibrium had been reached. Scans at 360 nm were collected and subtracted from the equilibrium scans to correct for anomalies arising from the cell windows. The resulting *A* versus r data were analysed using the software supplied with the instrument (XLAEQ and XLAMW), employing the sedimentation equilibrium equation (1), where R is the gas constant, T is the absolute temperature, ω is the angular rotation in radians/sec, $\bar{v}$ is the partial specific volume (assumed to be 0.73 mL/g in this case) and ρ is the density of the buffer. For a homogeneous, thermodynamically ideal sample, equation (1) gives the molecular weight from the slope of a plot of ln*A* versus r^2. If the sample is not homogeneous, ln*A* versus r^2 is a function of r (and hence concentration). Equation (1) then yields point weight average molecular weights, each being characteristic of the composition mixture produced by the centrifugal fraction at that radial distance in the cell at which the slope has been evaluated.

$$M = \frac{2RT}{(1-\bar{v}\rho)\omega^2} \cdot \frac{d\ln A}{d(r^2)} \qquad (1)$$

4.3.2 Reduction and Carboxymethylation of Proteins

Samples were reduced and carboxymethylated as follows. Protein samples (5 mg/mL) in 6 M guanidine.HCl, 0.1 M Tris.HCl, 1 mM EDTA were adjusted to pH 8.3 and a solution of 0.1 M DTT was added to give a final concentration of 2 mM. Samples were then sealed under nitrogen and treated at 38°C for 1 hour. A 50 mM solution of iodoacetic acid (neutralised with NaOH) was added to give a final concentration of 5

mM. Samples were again sealed under nitrogen and treated at 38°C in the dark for 1 h. The reaction was quenched with addition of 1% (v/v) β-mercaptoethanol. The mixture was then dialysed (5 changes) against milliQ water during which the reduced and carboxymethylated DnaA protein precipitated. The precipitate was collected by centrifugation, dried under vacuum, and analysed by ESI-MS (Section 4.3.5). ESI-MS was also used to analyse tryptic fragments of the reduced and carboxymethylated protein.

4.3.3 Tryptic Digestion of Proteins

Approximately 0.5 mg of protein, either native or reduced and carboxymethylated, was resuspended in 50 mM ammonium bicarbonate pH 7.8 (0.4 mL), and 1% (w/w) of bovine pancreatic, TPCK-treated trypsin (Promega) in 50 mM acetic acid was added. Samples were treated at 30°C overnight. Insoluble material was removed by centrifugation and samples were stored at -20°C until required.

4.3.4 HPLC Purification of Tryptic Peptides

Tryptic peptides were separated by high performance liquid chromatography (HPLC). HPLC was carried out using a Waters Millipore HPLC system to which a Rheodyne injector had been attached. The system consisted of two model 451 HPLC pumps, a model 680 automated gradient controller and a model 441 absorbance detector with a 214 nm filter. The column employed was an Alltech alltima 5 micron C18 reverse phase HPLC column (250 x 4.6 mm). It was equilibrated with 10% aqueous acetonitrile containing 0.1% TFA (solvent A). Separation was achieved by a gradient elution program consisting of 5 min at 0% solvent B (neat acetonitrile, 0.1% TFA) followed by a linear gradient from 0-55% solvent B over 60 min at a flow rate of 1 mL/min. All peak fractions were collected and stored at -20°C until required.

4.3.5 Electrospray Mass Spectrometry (ESI-MS)

Protein samples for ESI-MS were dialysed (3 changes) against milliQ water to desalt the protein followed by dialysis (2 changes) against 0.1% formic acid to solubilise precipitated protein. The proteins were freeze dried using a Speed-vac and stored at -20˚C until required. A portion of the sample was resuspended to ~4 μM in 0.1% formic acid and 20 μL was used for ESI-MS.

Peptides generated by treatment with trypsin were analysed directly from the HPLC solvent at the composition in which they eluted from the column. The concentration of the tryptic peptide solutions were of the order of 1 μM.

All electrospray mass spectra were acquired on a VG Quattro II mass spectrometer (VG Biotech Ltd.) equipped with an electrospray ionisation source and a quadrupole-hexapole-quadrupole mass analyser configuration with a 4000 m/z range for singly charged species. The solvent (50% aqueous MeOH) for the spray was delivered by a HP 1090 liquid chromatograph at a flow rate of 8 μL/ min. A nitrogen nebulising gas, flowing concurrent to the stainless steel capillary in the probe at a flow rate of ~10 L/h, assisted in the production of a stable spray. A dry, warm, counter-current nitrogen bath gas at a flow rate of ~350 L/h and at atmospheric pressure was employed to assist evaporation of the electrospray solvent. The electrospray probe tip potential was 3.2 kV with 0.2 kV on the pepper-pot counter electrode. Skimmer potentials ranging from 45-65 V were used, and the resolution was set at 0.5 Da to distinguish doubly- and triply-charged ions. The photomultiplier was set at 650 V. All mass spectra were acquired in the positive ion mode by multi-channel analysis (MCA) at a rate of 1 s per m/z 100. Typically, data from 10-20 scans were summed to obtain the representative spectra. All mass spectra acquired were processed using the software provided by VG Biotech. The mass spectrophotometer was calibrated using Cs/RbI for analysis of tryptic peptides and horse heart myoglobin (M_r 16951.5, 4 μM) for analysis of intact

protein.

4.3.6 Tandem Mass Spectrometry (ESI-MS/MS)

For MS-MS experiments, argon was used as the collision gas at a pressure corresponding to 30% transmission of the incident ion beam. The collision energy varied from 0-40 eV as indicated. The photomultiplier was set at 850 V for the second mass analyser and the resolution set at 1 Da. Data collection and processing of mass spectra were performed as described for ESI-MS.

4.3.7 Data Processing Parameters for ESI-MS

All electrospray mass spectra presented in this thesis were processed using the Masslynx software supplied with the mass spectrometer.

Typically for ESI-MS experiments, the low and high mass resolution parameters (denoted LM res and HM res respectively) were both set at 16 arbitrary units, which translates to a resolution of approximately 0.3-0.5 Da peak width at half height. The resulting raw data were then baseline subtracted with either a linear subtraction or by one of a range of fitted polynomial subtractions supplied in the software. The subtracted spectrum was then smoothed 2 times half the peak width at half peak height (i.e., if the resolution is set at 1.0 Da peak width-at-half height, then the spectrum would be smoothed 2 x 0.45-0.55 Da) according to the 'mean' mathematical smoothing method. The smoothed MCA data were then converted to a peak-top spectrum (stick spectrum) upon which the true mass was calculated. The true mass conversion (transformed spectrum) was carried out on the subtracted raw data either by a simple mathematical transformation or by a complex maximum entropy deconvolution method (Ferrige *et al.*, 1991) (Maxent™). The transformed spectra were smoothed again 2 times half the peak width-at-half height (although the peak width was generally larger

for the transformed spectra than for the raw data).

4.3.8 In vitro Phosphorylation of DnaA

For *in vitro* phosphorylation experiments, DnaA protein (400 pmol) was added to a solution of 1 M Tris.HCl pH 7.5, 2 M potassium glutamate, 0.2 M magnesium acetate and 5 mM γ-[^{32}P] ATP (~1 mCi). Crude cell extract (from strain WM433, Fraction II) or protein (DnaK or GroES-EL) were added in varying concentrations. The solutions were then mixed and Treated at 30°C for 10 min. Control reactions contained no DnaA. Following treatment, 20 µL of SDS-loading buffer was added and the samples were loaded onto a 15% SDS-PAGE gel. After electrophoresis, proteins were stained with Coomassie brilliant blue, then the gel was placed on Whatman 3MM paper, dried with a BioRad model 483 gel drier, then exposed to X-ray film (Kodak XAR-5) with intensifying screen (Dupont) at -70°C for 7 days. The autoradiograph was analysed by comparison with the Coomassie blue stained gel.

4.3.9 UV Spectroscopy of Proteins

Scans were performed on a Cary 1E high performance UV-vis spectrophotometer (Varian). Samples of DnaA protein (~ 1 mg/mL) which had been exhaustively dialysed into 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$, 0.1 mM ATP, 1 M KCl at 4°C were scanned from 240-360 nm at 40 nm/sec using a slit width of 2 nm and a signal averaging time of 0.1 sec. Data were collected at 1 nm intervals. A scan of the external dialysis buffer alone provided the data for baseline correction.

4.3.10 *Simulation of the Protein Spectra using N-acetyl-L-tryptophan ethyl ester, N-acetyl-L-tyrosine ethyl ester and ATP*

Solutions of N-acetyl-L-tryptophan ethyl ester, N-acetyl-L-tyrosine ethyl ester and ATP (all from Sigma Chemical Co.) in 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$ and 1 M KCl were scanned and baseline corrected as described in Section 4.3.9.

Data were collected from the scans at 2 nm intervals from 240-290 nm and 5 nm intervals from 290-360 nm. Using EXCEL spreadsheet software (Microsoft), the data were adjusted so that the 280 nm value was equivalent to extinction coefficients of tryptophan (ε_{280} 5690 M^{-1} cm^{-1}) and tyrosine (ε_{280} 1280 M^{-1} cm^{-1}) (Gill and von Hippel, 1989). ATP data were adjusted similarly for a 260 nm value (ε_{260} 15 400 M^{-1} cm^{-1}). Equation (2) was used to obtain data to simulate spectra of a protein, where $\varepsilon_{p,\lambda}$, $\varepsilon_{trp,\lambda}$, $\varepsilon_{tyr,\lambda}$ and $\varepsilon_{ATP,\lambda}$ are the molar absorptivities of the protein, tryptophan, tyrosine and ATP, respectively, at wavelength λ. The coefficients *n* and *m* were the number of tryptophan and tyrosine residues per protein molecule, and *k* is the number of ATP molecules per protein molecule (set at either 0, 1 or 2). For DnaA, $n = 5$ and $m = 11$ (Hansen *et al.*, 1982). *E. coli* prolidase was used as a control and has a value of 5 for *n* and 15 for *m*.

$$\varepsilon_{p,\lambda} = n\varepsilon_{trp,\lambda} + m\varepsilon_{tyr,\lambda} + k\varepsilon_{ATP,\lambda} \qquad (2)$$

For comparison of the observed and predicted spectra, data were normalised at 280 nm to an absorbance of 1.0. Spectra were generated using KaleidaGraph.

4.4 Results and Discussion

4.4.1 Sedimentation Equilibrium Experiments with DnaA Protein

DnaA protein purified by Fuller and Kornberg (1983) and Sekimizu *et al.* (1988) was reported to exist in two forms, monomeric (active) and aggregated (inactive). The DnaA purified by the method described in Section 3.3.3 eluted in a single symmetrical peak from a Sephadex G-200 column at a position corresponding to a globular protein with a molecular weight of 77 kDa which suggested that the protein was monomeric but had a somewhat asymmetric shape. Sedimentation equilibrium studies were performed to further confirm that the protein was monomeric.

A solution of DnaA in buffer containing 25 mM MOPS pH 6.8, 0.1 mM EDTA, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$, 0.1 mM ATP and 1 M KCl was used for the experiments with the buffer as a reference. A molecular weight of 45.5 ± 0.3 kDa was obtained from a linear regression analysis of a plot of ln*A* versus r^2 (Figure 4.2), using Equation (1) (see Section 4.3.1). The results from sedimentation equilibrium experiments with DnaA have consistently given a molecular weight lower than the theoretical molecular weight of the monomeric species (52 419.7 Da). DnaA is known to bind phospholipids which may increase the partial specific volume ($\bar{v}$) above the assumed value of 0.73 cm^3/g and in turn affect the calculated result. This may account for the ~10% lower value for the molecular weight compared to the theroetical molecular weight based on the amino acid sequence. An extended time of 24 hours was required for equilibrium to be reached, which may suggest that the protein is asymmetric in shape. This was supported by results from Sephadex G-200 gel filtration chromatography (Section 3.4.4).

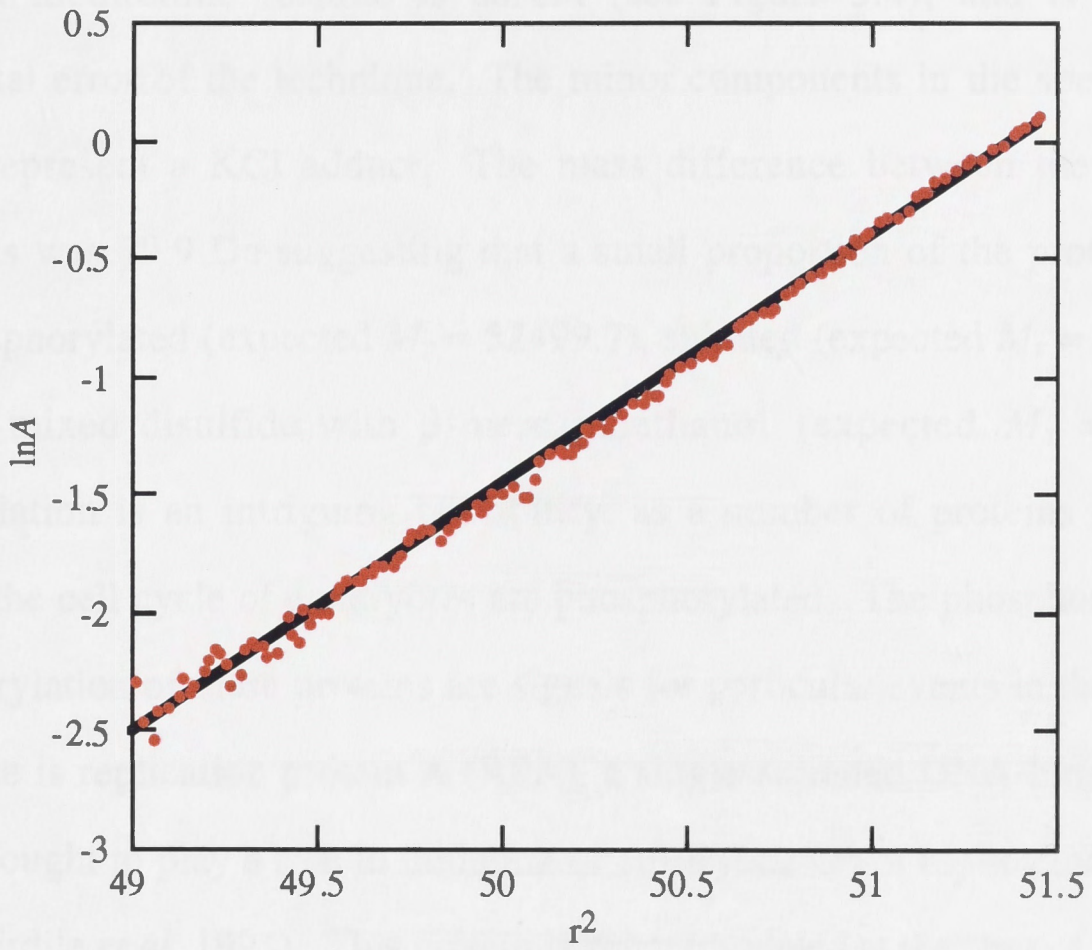


Figure 4.2

Sedimentation equilibrium analysis of DnaA represented by a plot of ln*A* versus r^2, where *A* is the absorbance and r is the radial distance from the axis of rotation. A molecular weight of 45.5 ± 0.3 kDa was calculated using linear regression analysis and equation (1) (Section 4.3.1).

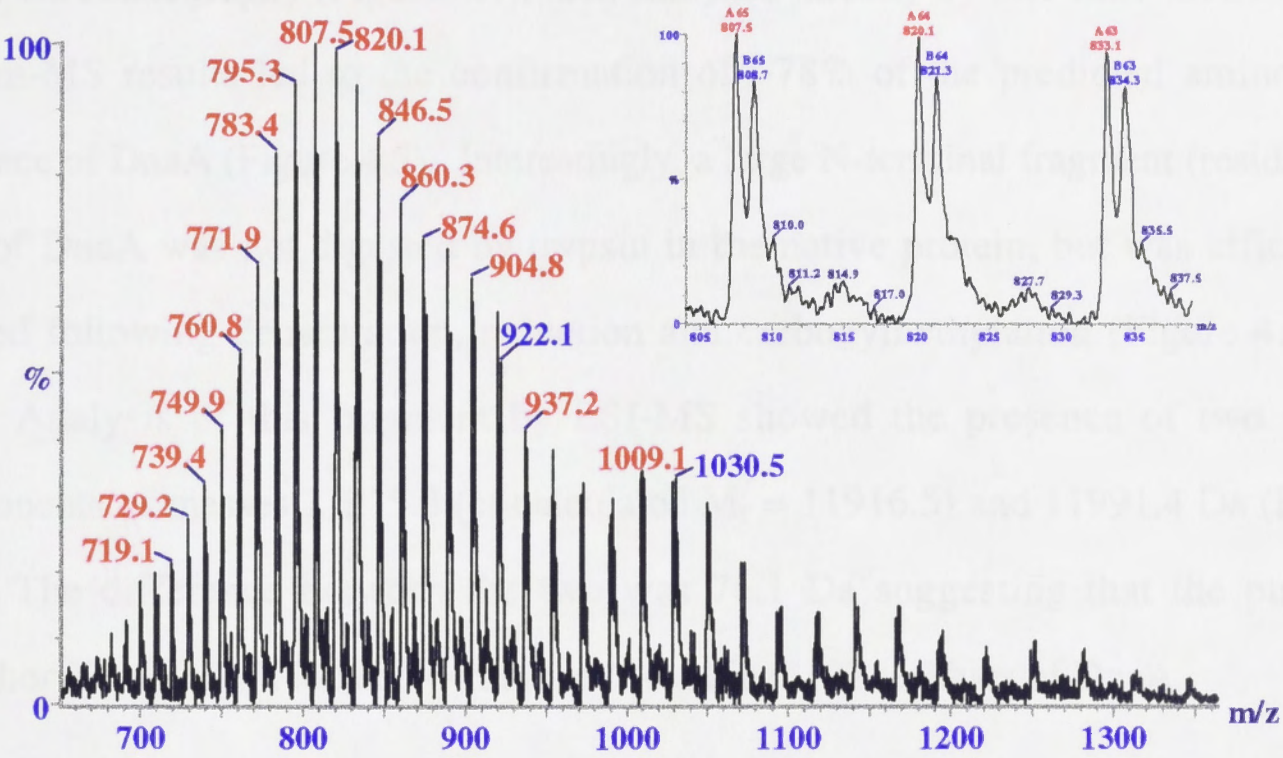
4.4.2 Electrospray Ionisation Mass Spectrometry of DnaA

The ESI-MS analysis of DnaA (prepared as described in Section 4.3.5) resulted in the identification of two major components of masses 52 422.5 and 52 500.4 Da and a minor component at 52 459.2 Da (Figure 4.3). The smaller of these masses (52 422.5 Da) is only 2.8 Da larger than the theoretical mass of DnaA (52 419.7 Da) assuming the N-terminal methionine residue is absent (see Figure 3.4), and is within the experimental error of the technique. The minor components in the spectrum most probably represent a KCl adduct. The mass difference between the two major components was 79.9 Da suggesting that a small proportion of the protein may be either phosphorylated (expected M_r = 52499.7), sulfated (expected M_r = 52499.7) or exist as a mixed disulfide with β-mercaptoethanol (expected M_r = 52496.8). Phosphorylation is an intriguing possibility, as a number of proteins involved in control of the cell cycle of eukaryotes are phosphorylated. The phosphorylation and dephosphorylation of these proteins are signals for particular events in the cell cycle. An example is replication protein A (RPA), a single-stranded DNA-binding protein which is thought to play a role in initiation of eukaryotic DNA replication (Wobbe *et al.*, 1987; Erdile *et al.*,1991). This protein is phosphorylated at the start of S phase in a cell cycle-dependent manner within the initiation complex by a cyclin-dependent kinase (*cdc2* kinase) (Dutta *et al.*, 1991; Dutta and Stillman, 1992; Fotedar and Roberts, 1992). De-phosphorylation occurs at mitosis to reset the phosphorylation cycle for nucleation of new replication complexes (Dutta *et al.*, 1991). Thus, it may be that phosphorylation of DnaA may play a role in regulation of *E. coli* chromosomal DNA replication.

4.4.3 Tryptic Digestion Analysis of DnaA

In order to identify the region of the protein that may be phosphorylated, the protein was digested with trypsin (as described in Section 4.3.3). A tryptic digest of the native

(a)



(b)

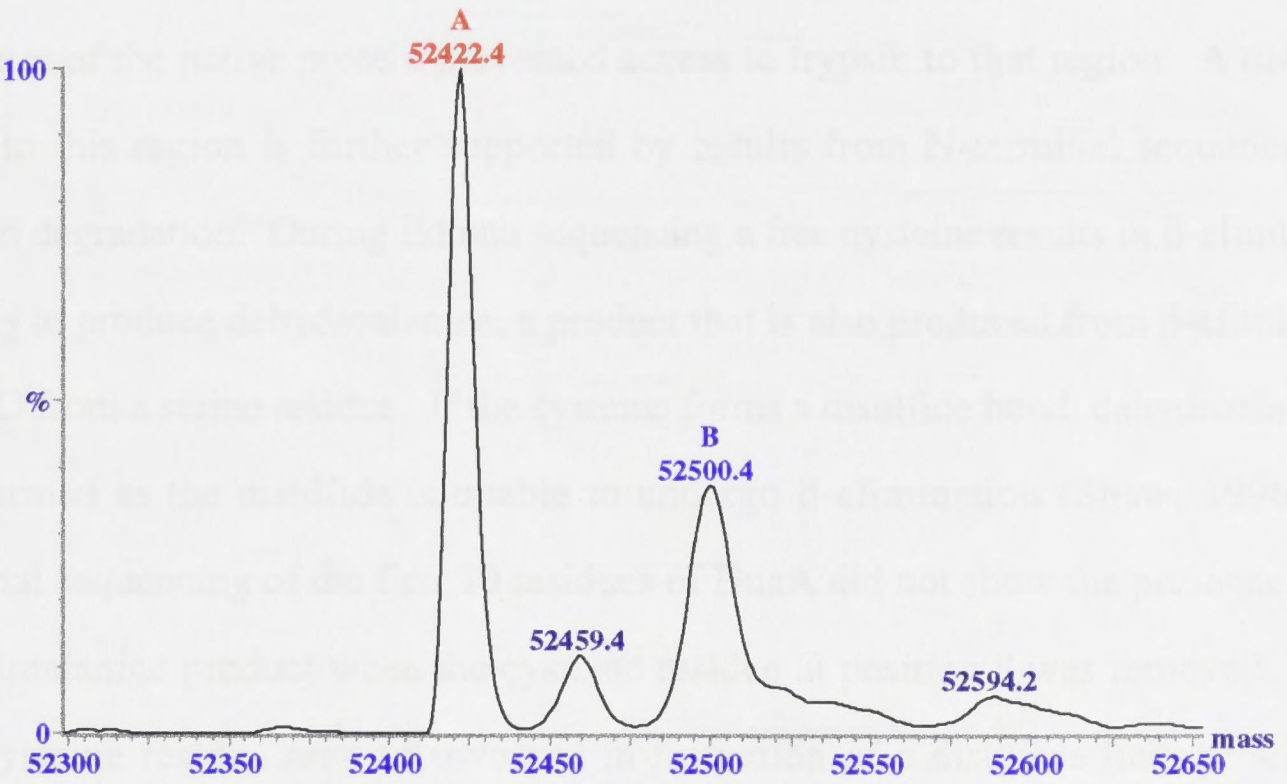


Figure 4.3

(a) A m/z spectrum (baseline subtracted and smoothed) of DnaA (calculated M_r = 52 419.7 based on the amino acid composition) obtained from ESI-MS at a skimmer potential of 50 V. The protein was diluted in an aqueous solution of 0.1% formic acid and 50% methanol.

(b) The transformed ESI-MS spectrum of DnaA generated using Maxent ™.

protein produced many peptide fragments that were separated by C18 reverse phase HPLC chromatography (Figure 4.4), then analysed directly by ESI-MS. ESI-MS and tandem-MS results led to the confirmation of ~78% of the predicted amino acid sequence of DnaA (Figure 4.5). Interestingly, a large N-terminal fragment (residues 1-107) of DnaA was not digested by trypsin in the native protein, but was efficiently cleaved following denaturation, reduction and carboxymethylation (Figure 4.4 and 4.5). Analysis of this fragment by ESI-MS showed the presence of two major components of masses 11915.3 (cf calculated M_r = 11916.5) and 11991.4 Da (Figure 4.6). The difference between the two was 76.1 Da suggesting that the putative phosphorylation site is located within the N-terminal 107 residues of DnaA.

It may be that this N-terminal peptide of DnaA was held intact in the native protein by a disulfide linkage between the two cysteine residues (positions 8 and 66) or that the structure of the native protein prevented access to trypsin to that region. A disulfide bond in this region is further supported by results from N-terminal sequencing by Edman degradation. During Edman sequencing a free cysteine results in β-elimination of SH_2 to produce dehydroalanine, a product that is also produced from β-elimination of H_2O from a serine residue. If the cysteine forms a disulfide bond, dehydroalanine is not formed as the disulfide is unable to undergo β-elimination (Shaw, 1996). N-terminal sequencing of the first 10 residues of DnaA did not show the presence of the dehydroalanine product when the cysteine residue at position 8 was removed. Thus, this cysteine residue may be involved in formation of a disulfide linkage with the cysteine residue at position 66, thus protecting the N-terminal 107 residues from trypsin digestion.

DnaA was reduced and carboxymethylated to facilitate the digestion of the N-terminal 107 residues by trypsin in an attempt to further pinpoint the position of the putative phosphorylation. The protein was reduced and carboxymethylated as described (Section 4.3.2), then digested with trypsin (Section 4.3.3) and the peptide fragments

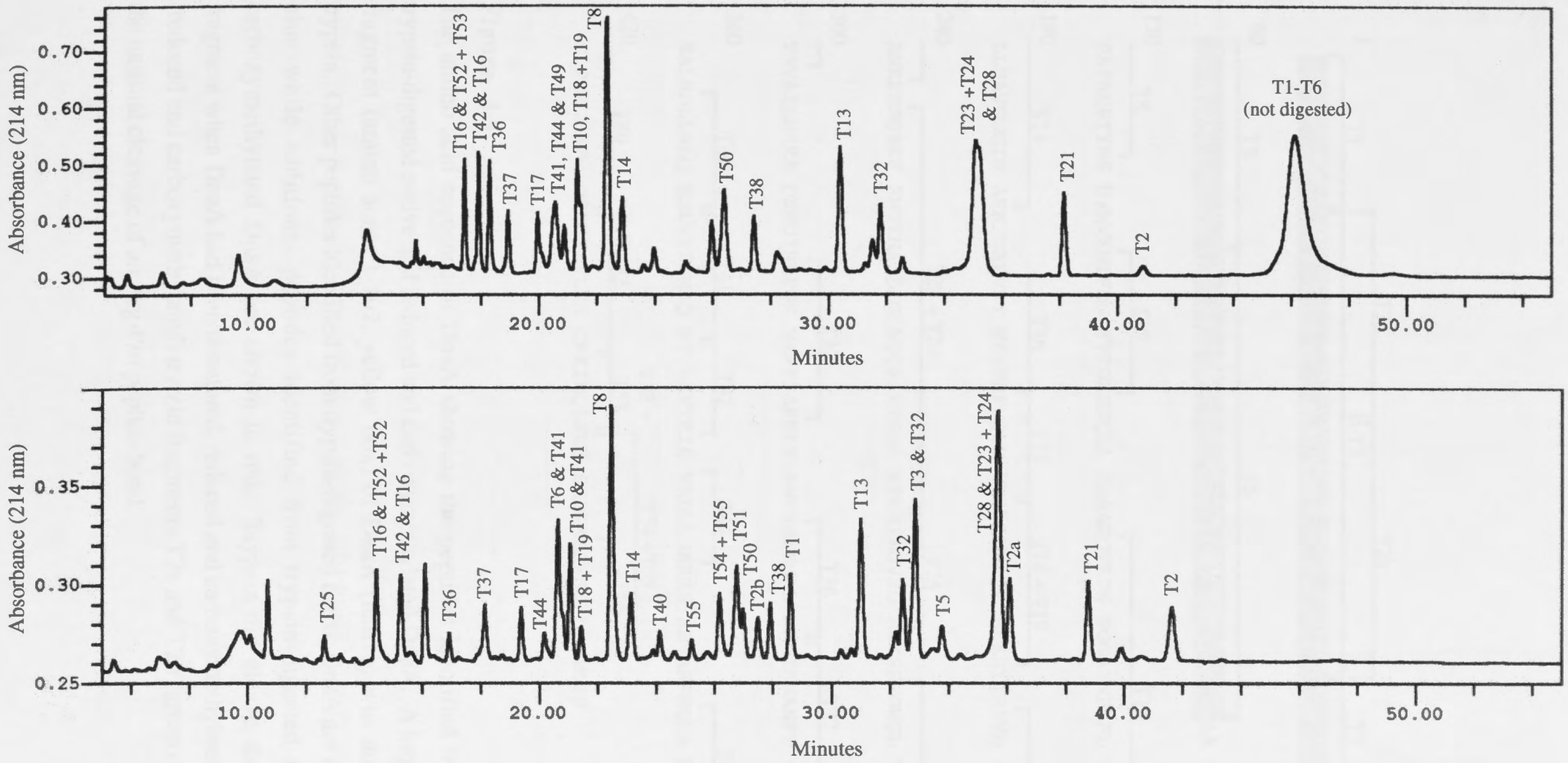


Figure 4.4

C18 reverse phase HPLC chromatograms of peptide fragments from native DnaA (top) and reduced and carboxymethylated DnaA (bottom) generated by digestion with trypsin. Tryptic peptides are numbered from the N-terminus and labelled with a "T" (see Figure 4.5).

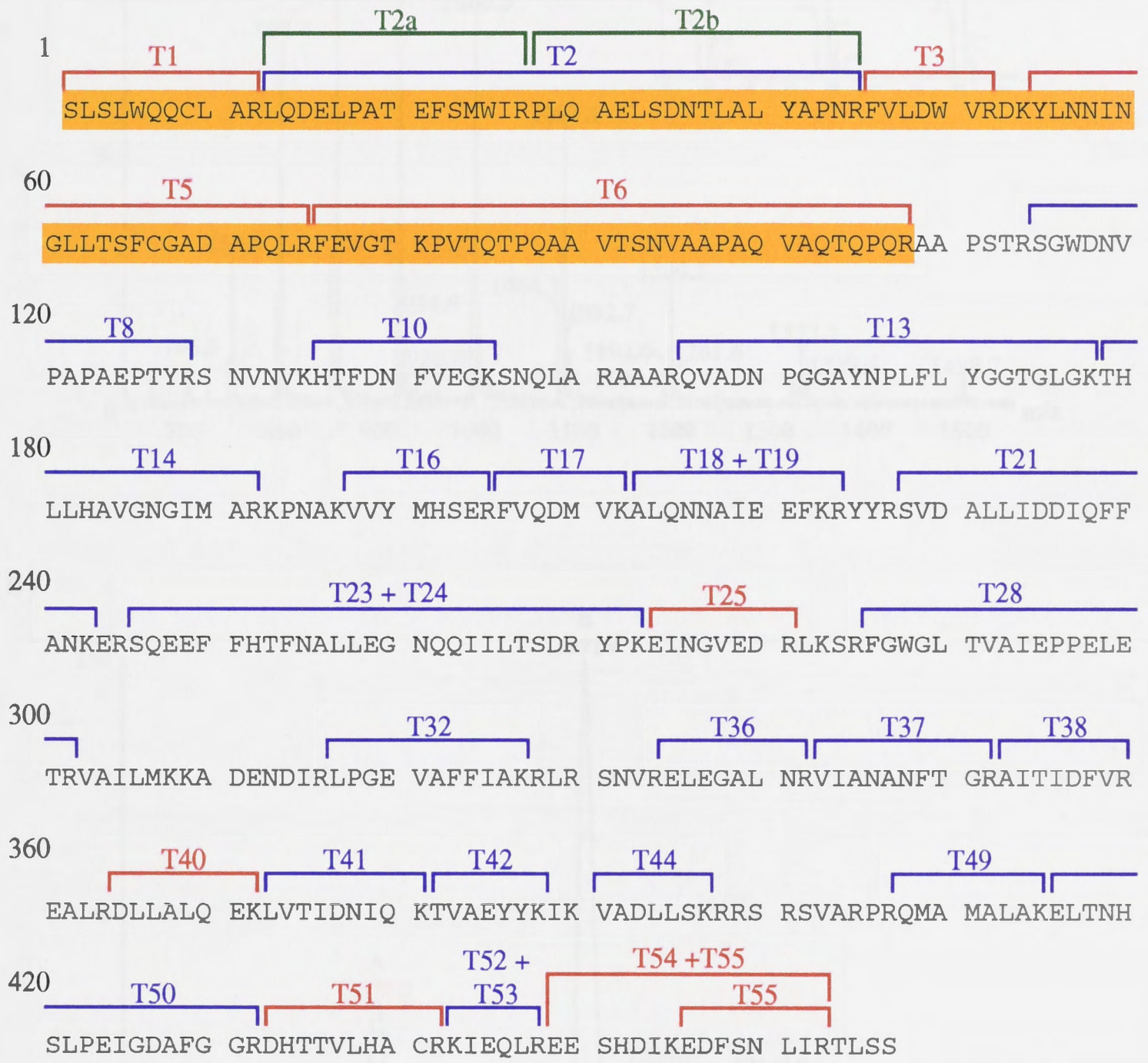


Figure 4.5

The amino acid sequence of DnaA showing the peptides identified by ESI-MS of trypsin-digested native and reduced and carboxymethylated DnaA. A large N-terminal fragment (amino acids 1-107, yellow box) of DnaA (native) was not cleaved by trypsin. Other peptides identified from trypsin-digested native DnaA are represented in blue while additional peptides identified from trypsin-digested reduced and carboxymethylated DnaA are shown in red. Trypsin did cleave the N-terminal fragment when DnaA had been denatured, reduced and carboxymethylated. The DnaA (reduced and carboxymethylated) tryptic fragments T2a and T2b (green) resulted from the unusual cleavage of an Arg-Pro peptide bond.

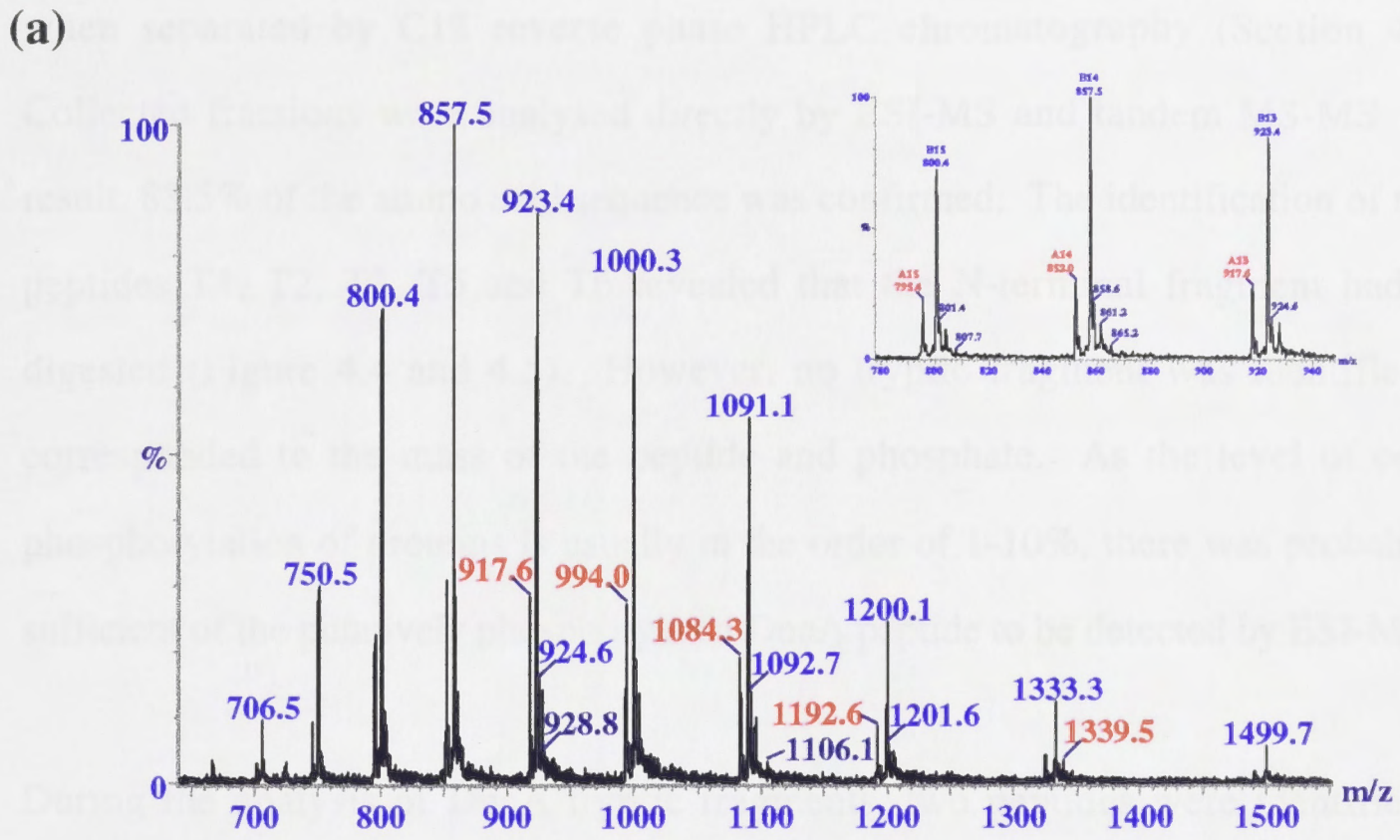


Figure 4.6

(a) A m/z spectrum (baseline subtracted and smoothed) of a tryptic peptide of DnaA (residues 1-107) obtained from ESI-MS at a skimmer potential of 50 V. The protein was analysed directly from a sample collected from C18 reverse phase chromatography.

(b) The transformed ESI-MS spectrum of DnaA tryptic peptide (residues 1-107) generated using Maxent ™. The M_r of this fragment was calculated to be 11916.5.

when separated by C18 reverse phase HPLC chromatography (Section 4.3.4). Collected fractions were analysed directly by ESI-MS and tandem MS-MS. As a result, 83.5% of the amino acid sequence was confirmed. The identification of tryptic peptides T1, T2, T3, T5 and T6 revealed that the N-terminal fragment had been digested (Figure 4.4 and 4.5). However, no tryptic fragment was identified that corresponded to the mass of the peptide and phosphate. As the level of cellular phosphorylation of proteins is usually in the order of 1-10%, there was probably not sufficient of the putatively phosphorylated DnaA peptide to be detected by ESI-MS.

During the analysis of DnaA tryptic fragments, two peptides were identified that resulted from the cleavage of an Arg-Pro bond. Reduced and carboxymethylated DnaA was digested with trypsin and the resulting peptides were identified by ESI-MS and tandem MS-MS. Peptides were identified that corresponded to tryptic fragment T2 (M_r 3803.16, Figure 4.7) as well as fragments T2a (M_r 1834.26) and T2b (M_r 1984.43) which were the result of an Arg-Pro bond cleavage (Figure 4.5). ESI-MS/MS was used to confirm the composition of peptides T2a and T2b (Figure 4.8). The fragment ions of the ESI-MS/MS spectra were assigned and are summarised in Tables 4.1 and 4.2. It is widely accepted that trypsin does not normally cleave Arg-Pro bonds. However, the detection of these peptides by ESI-MS suggests that Arg-Pro bonds may be cleaved by trypsin under certain conditions.

4.4 Phosphorylation of DnaA

In the previous section it was speculated that a portion of purified DnaA may be phosphorylated. This implies existence in the cell of a DnaA-kinase that performs the phosphorylation. Several *in vitro* experiments were set up in an attempt to phosphorylate DnaA with ^{32}P using the cell extract from a mutant *dnaA*-5 strain (WM433), DnaK (known to associate with DnaA) or the molecular chaperones GroES-EL. The reactions were performed as described in Section 4.3.8. An autoradiograph

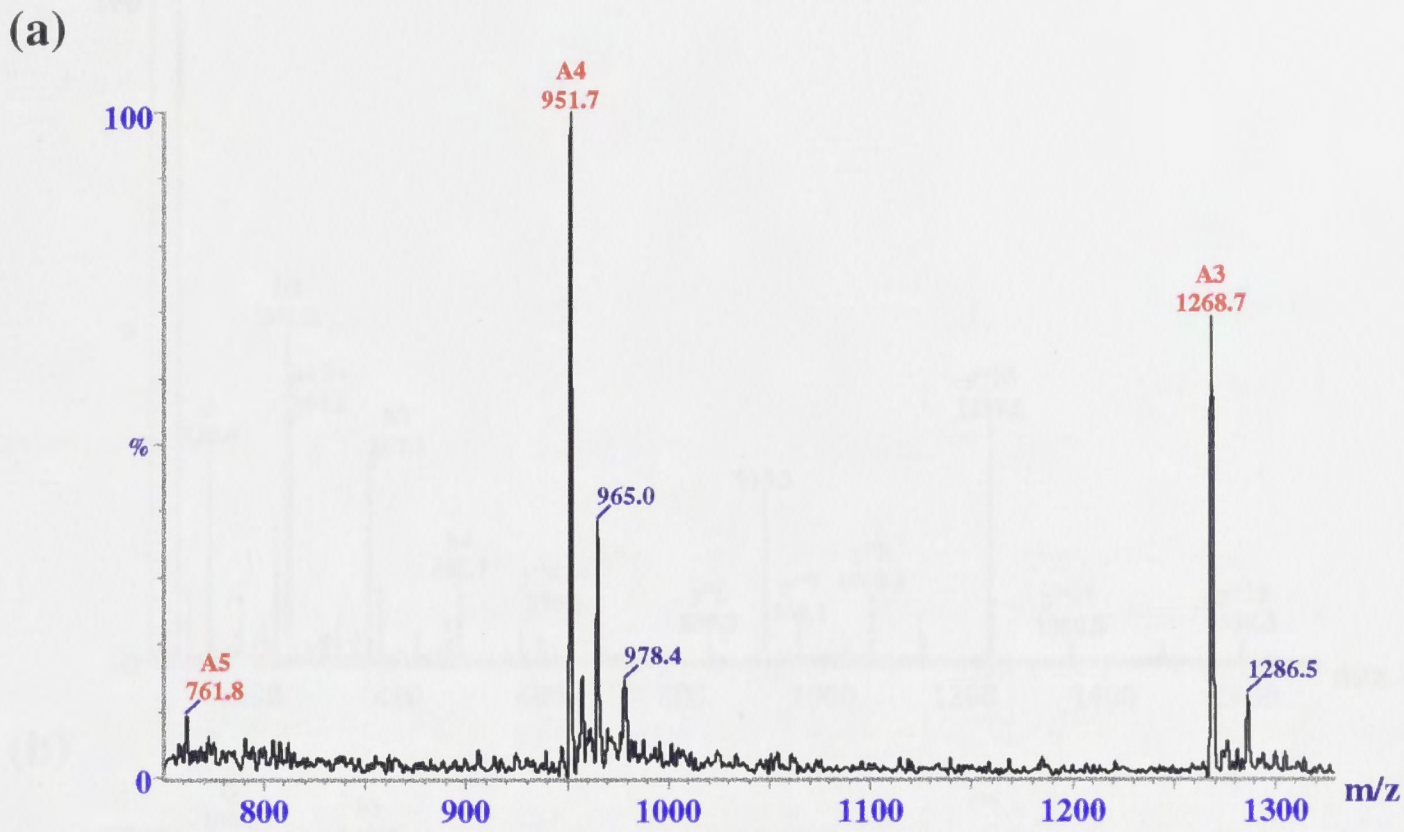
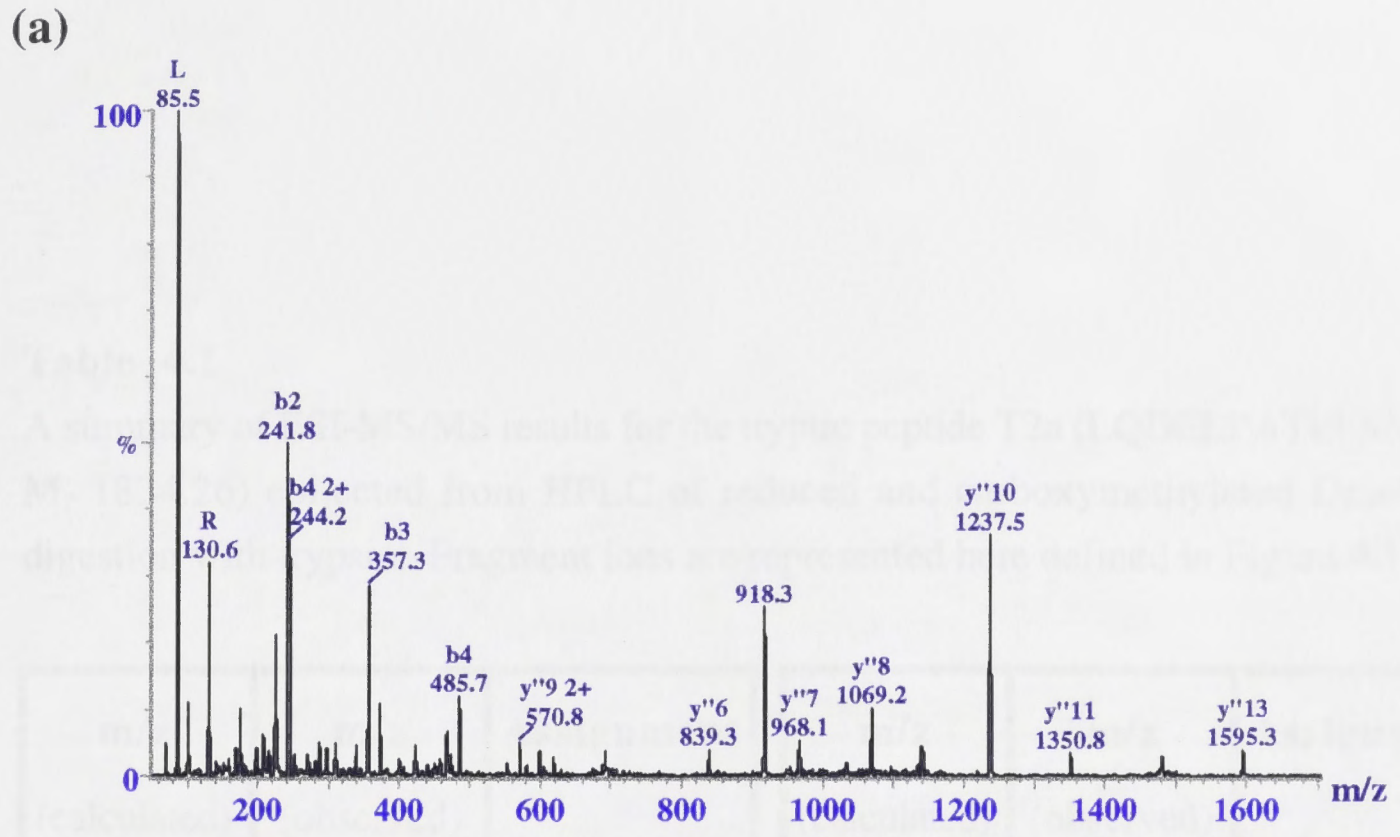


Figure 4.7

(a) A m/z spectrum (baseline subtracted and smoothed) of a tryptic peptide of DnaA (T2, residues 12-48, calculated M_r = 3804.3) obtained from ESI-MS at a skimmer potential of 50 V. The protein was analysed directly from a sample collected from C18 reverse phase HPLC chromatography.

(b) The transformed ESI-MS spectrum of DnaA tryptic peptide T2 (residues 12-48) generated using Maxent ™.



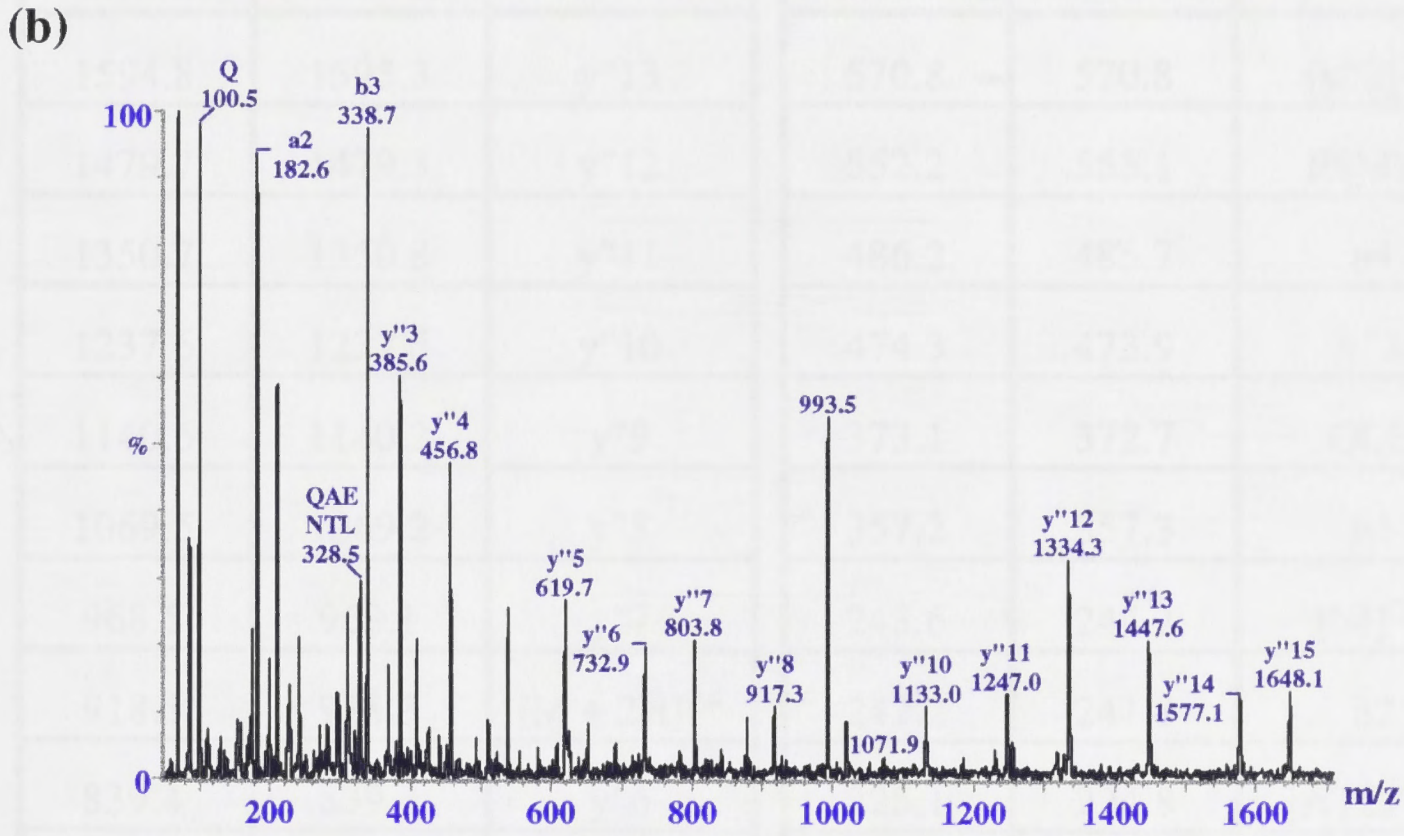


Figure 4.8

(a) A tandem spectrum (baseline subtracted and smoothed) of the doubly charged molecular ion (918.3) of the tryptic peptide T2a (residues 12-26, calculated M_r 1834.9) from reduced and carboxymethylated DnaA treated with trypsin. The tryptic peptide was analysed directly from a sample collected from C18 reverse phase HPLC chromatography.

(b) A tandem spectrum (baseline subtracted and smoothed) of the doubly charged molecular ion (993.5) of the tryptic peptide T2a (residues 12-26, calculated M_r 1985.0) from reduced and carboxymethylated DnaA treated with trypsin. The tryptic peptide was analysed directly from a sample collected from C18 reverse phase HPLC chromatography.

Table 4.1

A summary of ESI-MS/MS results for the tryptic peptide T2a (LQDELPATEFSMWIR, M_r 1834.26) collected from HPLC of reduced and carboxymethylated DnaA after digestion with trypsin. Fragment ions are represented here defined in Figure 4.1.

m/z (calculated)	**m/z** (observed)	**Assignment**	**m/z** (calculated)	**m/z** (observed)	**Assignment**
1594.8	1595.3	y"13	570.8	570.8	$[y"9]^{2+}$
1479.7	1479.3	y"12	552.2	553.1	FSMW
1350.7	1350.8	y"11	486.2	485.7	b4
1237.6	1237.5	y"10	474.3	473.9	y"3
1140.6	1140.2	y"9	373.1	372.7	QDE
1069.5	1069.2	y"8	357.2	357.3	b3
968.5	968.1	y"7	243.6	244.2	$[b4]^{2+}$
918.5	918.3	$[M + 2H]^{2+}$	242.2	241.8	b2
839.4	839.3	y"6	225.1	224.8	$[ATEF]^{2+}$
696.4	676.7	b7	175.1	174.4	y"1
692.4	692.2	y"5	169.1	168.6	PA
619.3	619.1	$[y"10]^{2+}$	101.1	100.5	Q
605.3	604.7	y"4	86.1	85.5	L
599.3	599.2	b5	70.1	69.3	P

Table 4.2

A summary of ESI-MS/MS results for the tryptic peptide T2b (PLQAELSDNTLALYAPNR, M_r 1984.43) collected from HPLC of reduced and carboxymethylated DnaA after digestion with trypsin. Fragment ions represented here are defined in Figure 4.1.

m/z (calculated)	m/z (observed)	Assignment
1647.8	1648.1	y"15
1576.8	1577.1	y"14
1447.8	1447.6	y"13
1334.7	1334.3	y"12
1247.6	1247.0	y"11
1132.6	1133.0	y"10
1018.6	1018.4	y"9
993.5	993.5	$[M+2H]^{2+}$
917.5	917.3	y"8
804.4	803.8	y"7
733.4	732.9	y"6
652.4	652.4	b6
924.3	624.2	$[y"11]^{2+}$
620.3	619.7	y"5
555.3	555.0	LQAEL
539.3	538.9	b5

m/z (calculated)	m/z (observed)	Assignment
509.8	510.2	$[y"9]^{2+}$
457.3	456.8	y"4
410.2	409.4	b4
386.2	385.6	y"3
339.2	338.7	b3
329.15 or 329.18	328.5	QAE or NTL
310.7	310.9	$[y"5]^{2+}$
242.2	241.8	LQ
209.6 or 210.1	210.5	$[SDNT]^{2+}$ or $[ALYA]^{2+}$
183.1	182.6	a2
175.1	174.7	y"1
101.1	100.5	Q
86.1	85.6	L
70.1	69.7	P
60.0	60.1	S

showed that several proteins in the extract were labelled with ^{32}P by the crude WM433 cell extract. One of these ^{32}P labelled proteins was in a position that corresponded to DnaA and appeared only in lanes to which purified DnaA was added (Figure 4.9). However, DnaA was not phosphorylated in experiments with DnaK and GroES-EL (data not shown). It is possible to speculate that some factor in the cell extract is required for phosphorylation of DnaA. Further studies are required to confirm that DnaA is phosphorylated and to determine the enzyme that performs this function.

4.5 Determination of the Number of ATP Molecules Bound to DnaA

The UV spectrum (240-360 nm) of DnaA (Figure 4.10) is not typical of a "normal" protein and this may be related to the binding of ATP. DnaA is thought to bind ATP at two sites, one at high affinity (K_d = 0.03 μM) and another at low affinity. Using the data from the spectra of ATP, N-acetyl-L-tryptophan ethyl ester and N-acetyl-L-tyrosine ethyl ester (Figure 4.11) spectra of DnaA could be simulated to contain zero, one or two molecules of ATP. These predicted spectra could be compared with the observed spectrum to determine the number of ATP molecules bound. However, before attempting the simulation it was necessary to compare the observed versus predicted spectrum of a protein which did not bind ATP to see if it were possible to simulate a protein spectrum. For this purpose a UV (240-360 nm) spectrum of *E. coli* prolidase (1 mg/mL, prepared by P.E. Lilley) was obtained as described in Section 4.3.9 and compared to the simulated spectrum derived by the method in Section 4.3.10. The predicted spectrum was very similar to the observed (Figure 4.12) except that the peak had shifted from 280 nm to approximately 275 nm. These results suggest that it is possible to approximately simulate the UV spectra of proteins.

DnaA spectra were simulated with zero, one and two molecules of ATP and compared with the observed DnaA spectrum (Figure 4.13). The predicted spectrum with 1 ATP molecule showed the most similarity to the observed spectrum although there was a

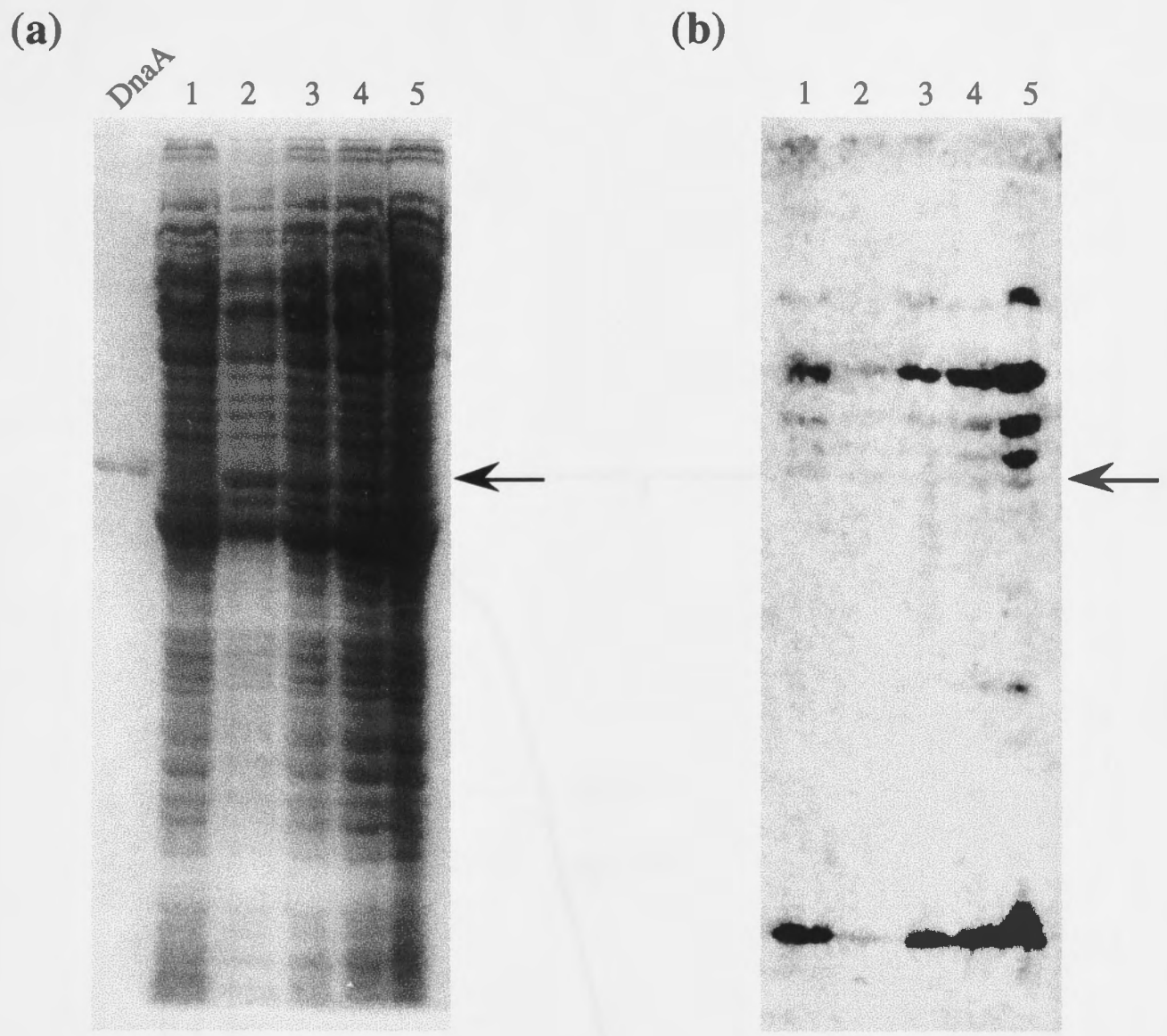


Figure 4.9

(a) SDS-PAGE of *in vitro* phosphorylated proteins. Samples were treated with cell extract (WM433, Fraction II) and γ-[^{32}P] ATP and a standard shows the position of DnaA on the gel. Cell extract only (lane 1) was used as a control while samples (lanes 2-5) contained DnaA (2 μg) and varying amounts of WM433 cell extract (0.2 μL, lane 2; 0.5 μL, lane 3; 1 μL, lane 4; 2 μL, lane 5). All samples contained 1 μCi of γ-[^{32}P]ATP. The mixtures were treated at 30°C for 10 min, then 20 μL of SDS loading buffer was added and samples applied to a 12% SDS-PAGE gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue. DnaA is indicated by the arrow.

(b) An autoradiograph of the SDS-gel (a) after exposure to X-ray film (XAR-5) indicating proteins that were labelled with ^{32}P. The arrow shows the position of DnaA.

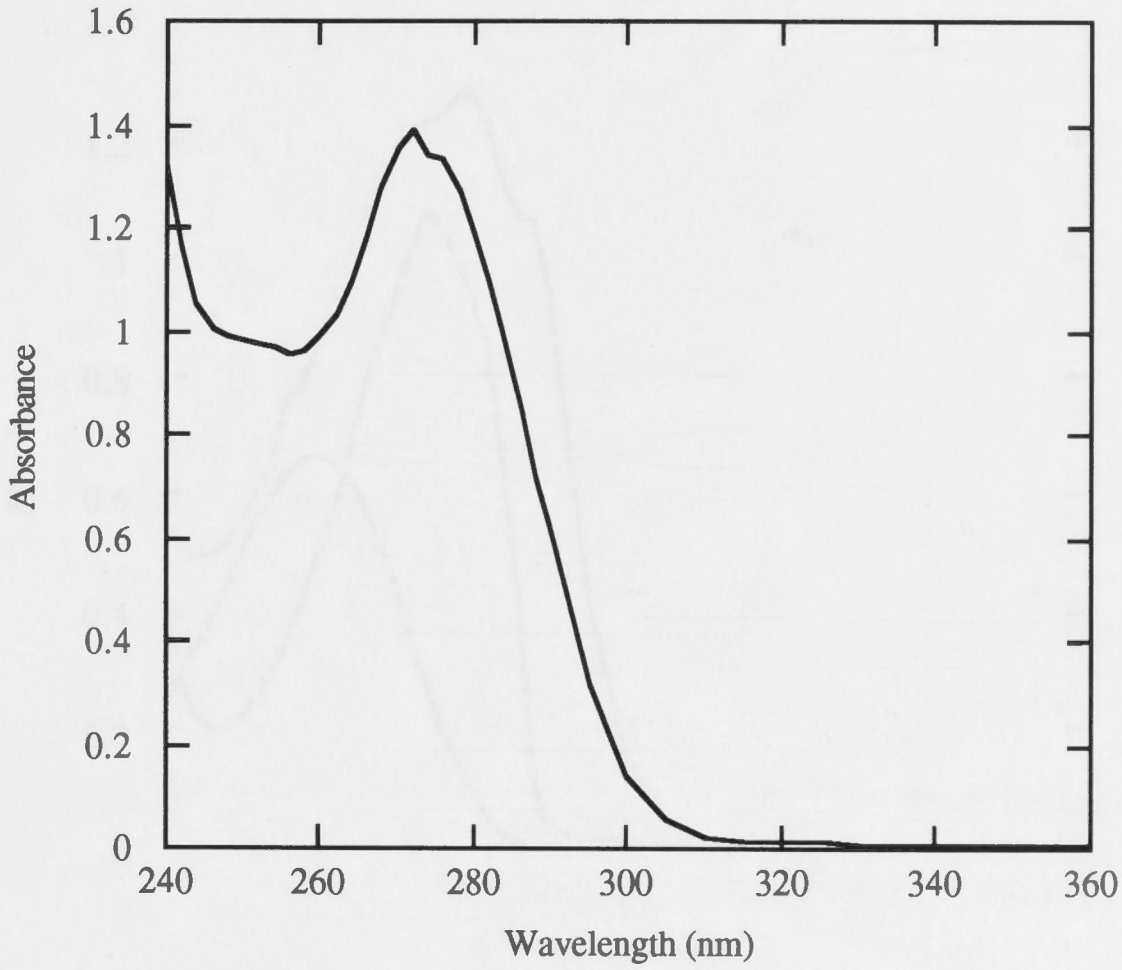


Figure 4.10

A UV spectrum of DnaA (~1 mg/mL) from 240-360 nm. Data were collected at 0.67 nm intervals with a signal averaging time of 0.1 sec, scanned at a rate of 40 nm/min and the solvent baseline was subtracted.

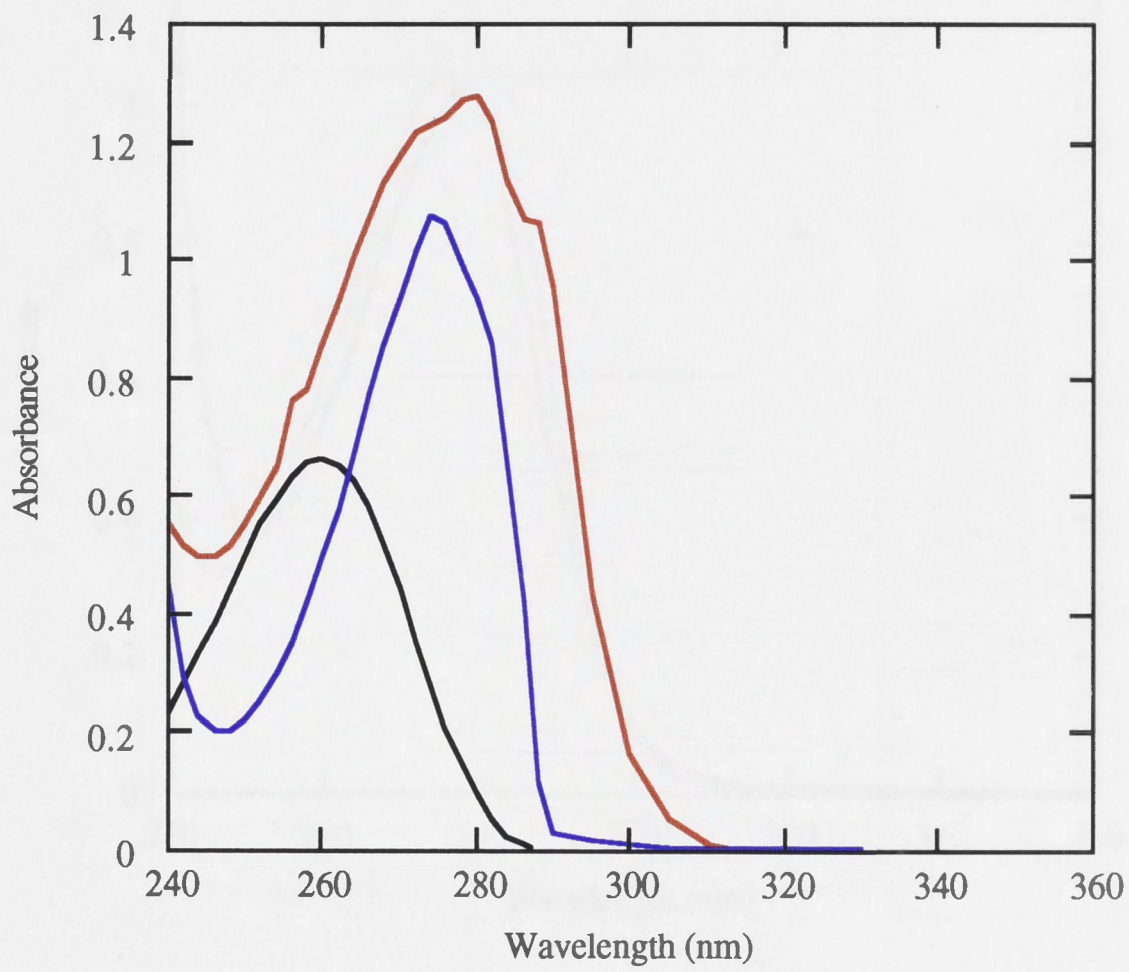


Figure 4.11

Spectrum of N-acetyl-L-tryptophan ethyl ester (~0.025 mg/mL, red), N-acetyl-L-tyrosine ethyl ester (~0.1 mg/mL, blue) and ATP (~0.02 mg/mL, black) from 240-360 nm. Data were collected at 0.67 nm intervals with a signal averaging time of 0.1 sec, scanned at a rate of 40 nm/min and the solvent baseline was subtracted.

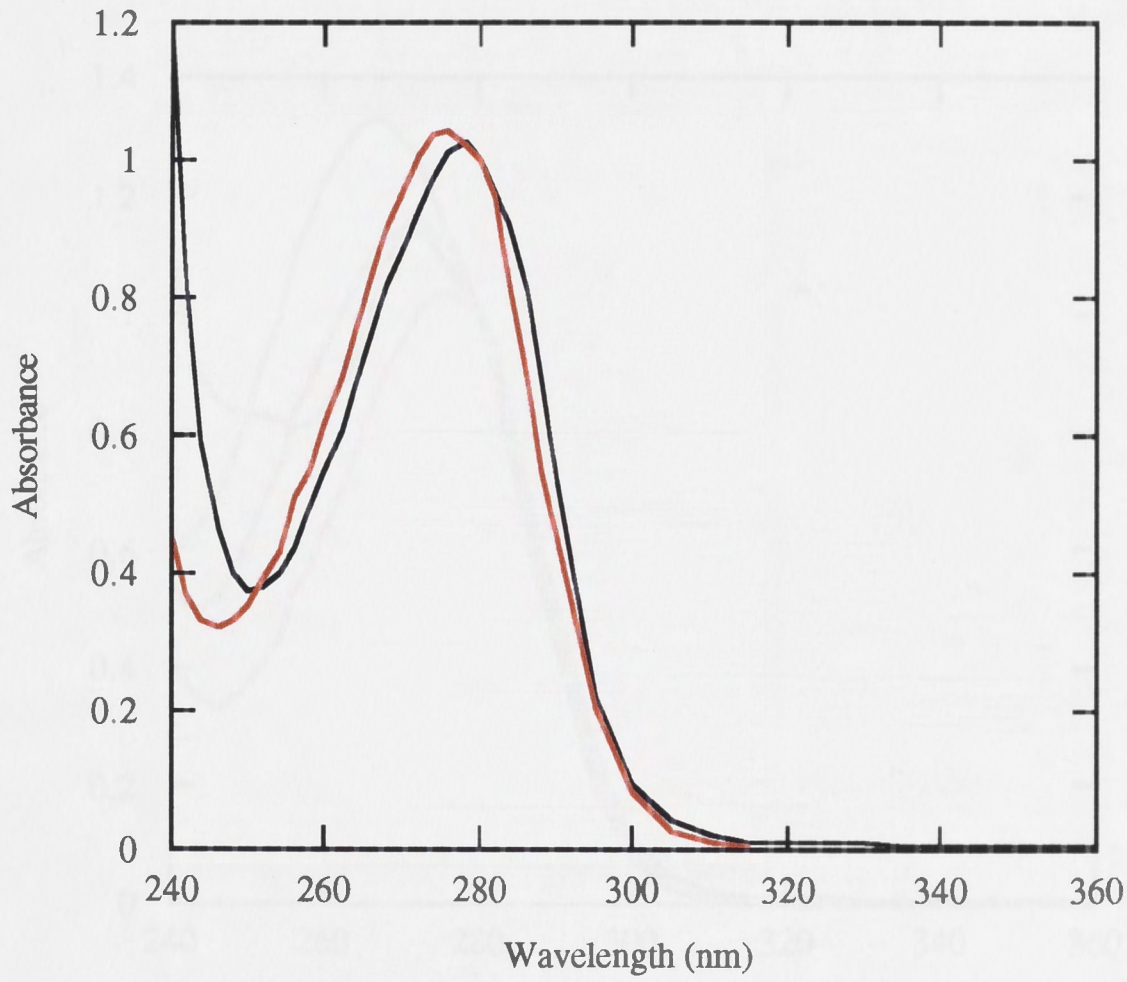


Figure 4.12

The spectrum of *E. coli* prolidase (~1 mg/mL, black) from 240-360 nm compared to the predicted spectrum (red) of the protein (each normalised to $A_{280} = 1.000$). Spectral data were collected at 0.67 nm intervals with a signal averaging time of 0.1 sec, scanned at a rate of 40 nm/min and the solvent baseline was subtracted. The spectra were then normalised for comparison with the predicted spectrum. The predicted spectrum was derived as described in Section 4.3.9, with $n = 5$, $m = 15$ and $k = 0$.

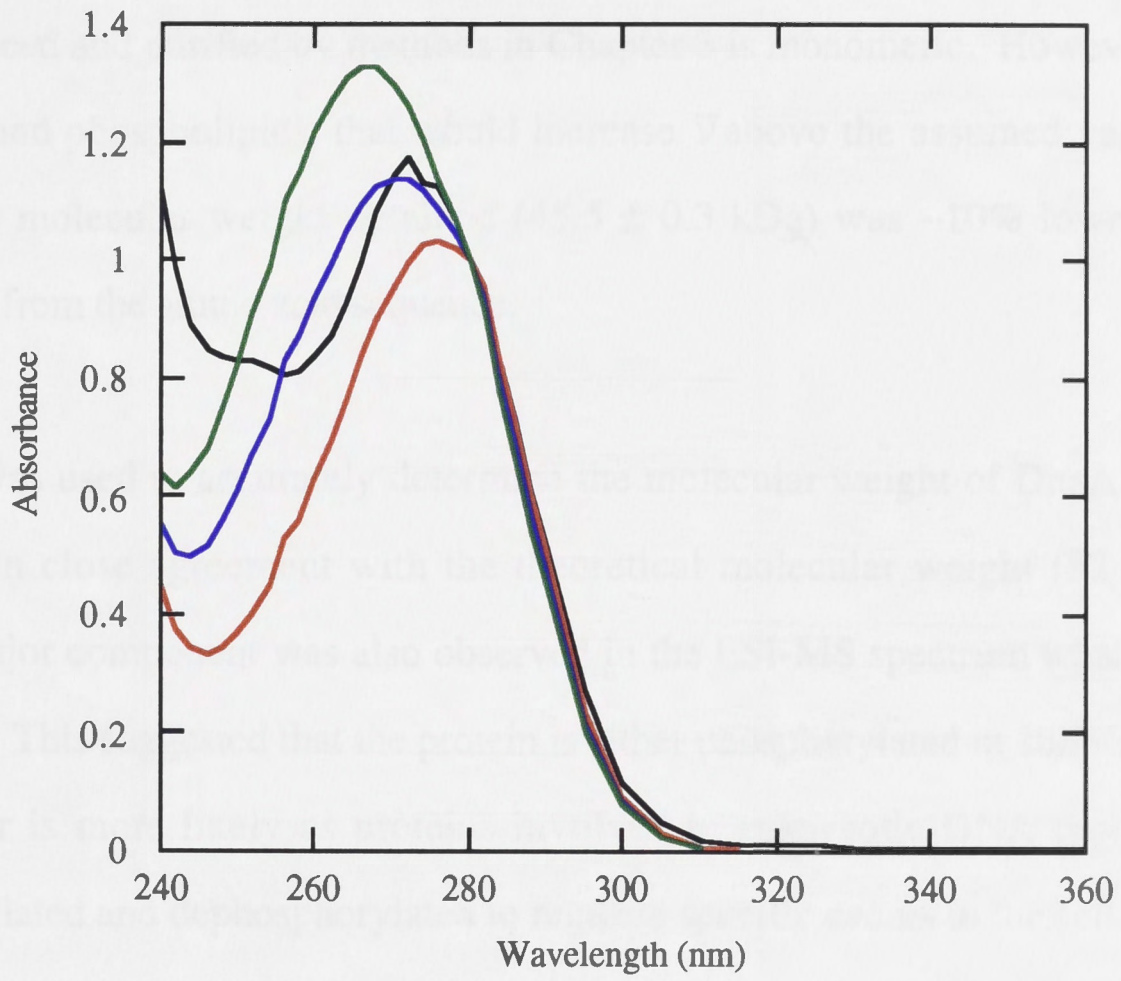


Figure 4.13

The spectrum of DnaA (normalised at 280 nm, black) from 240-360 nm compared with predicted spectra containing 0 (red), 1 (blue) or 2 (green) ATP molecules. The DnaA spectrum (Figure 4.9) was normalised as described in Section 4.3.9 and the predicted spectra generated using the method described in Section 4.3.9.

variation at 260 nm. Based on the comparison it appears as though there is absorbance at 260 nm in the observed DnaA spectrum which may be due to a chromophore other than ATP. However, it is also possible that the spectrum of bound ATP is very much different from that free in solution.

4.5 Conclusion

The results of sedimentation equilibrium experiments have shown that DnaA as overproduced and purified by methods in Chapter 3 is monomeric. However, perhaps due to bound phospholipids, that would increase $\bar{v}$ above the assumed value of 0.73 cm^3/g the molecular weight obtained (45.5 ± 0.3 kDa) was ~10% lower than that calculated from the amino acid sequence.

ESI-MS was used to accurately determine the molecular weight of DnaA (52 422.5) which is in close agreement with the theoretical molecular weight (52 419.7). A second major component was also observed in the ESI-MS spectrum which was 79.9 Da larger. This suggested that the protein is either phosphorylated or sulfated although the former is more likely as proteins involved in eukaryotic DNA replication are phosphorylated and dephosphorylated to regulate specific events in the cell cycle. It is possible that DnaA function is regulated in a similar manner in *E. coli* chromosomal DNA replication. Experiments to phosphorylate DnaA *in vitro* with ^{32}P using a crude cell extract from a *dnaA* mutant strain (WM433) produced labelled protein on an autoradiograph that corresponded to the position of DnaA. Although the results suggest that DnaA is phosphorylated, further studies are required to confirm these observations.

Digestion of DnaA with trypsin and analysis of the peptide fragments by ESI-MS and tandem MS/MS have shown that the site of putative phosphorylation is within the first 107 residues of DnaA. This region of the protein was found to contain a disulfide

bond between cysteine residues at positions 8 and 66 which prevents the native protein from being digested by trypsin. Determination of the site of phosphorylation is under active investigation at the time of writing this thesis.

Attempts were made to determine the number of ATP molecules bound to DnaA in solution by comparison of its UV spectrum with simulated protein spectra. Although it was possible to simulate the protein spectrum of *E. coli* prolidase, the spectrum of DnaA could not be simulated accurately. At best the comparison of the DnaA spectrum with spectra simulated with zero, one or two molecules of ATP would suggest that only one molecule of ATP is bound to DnaA. However, there appears to be absorbance at 260 nm in the DnaA spectrum from a chromophore other than ATP.

CHAPTER 5

CLONING, OVERPRODUCTION AND PURIFICATION OF M13 Gene II PROTEIN

5.1 Aims and Significance

The Gene II protein is the initiator of DNA replication of filamentous phages such as M13. The initial aim was to clone *gene II* into a suitable vector to improve expression of its product. Vectors for high-level overproduction include the λ promoter vectors (pPL450 series, Figure 2.1) developed in our laboratory and the pET vectors described by Studier *et al.* (1990). Improved levels of protein expression are vital for investigation of their functional intricacies.

To exploit the high-level overproduction of proteins, a sound method for their purification is essential. The current methods for purification of Gene II protein are laborious and the yields are extremely low (Meyer and Geider, 1979a; Greenstein and Horiuchi, 1987). Recently, the development of immobilised metal chelate affinity chromatograhy has become a widely used technique for the purification of recombinant proteins in a one-step procedure (Schmitt *et al.*, 1993). The resin most commonly used is Ni(II)-NTA (nitrilotriacetic acid) linked to beaded agarose. To utilise this purification method proteins require a poly-histidine tag at either the N- or C-terminus. The construction of vectors for tagging proteins with a poly-his tail are required and these vectors must also direct high-level overproduction of the proteins. Development of a new strategy for the purification of Gene II protein is the focus of work presented here.

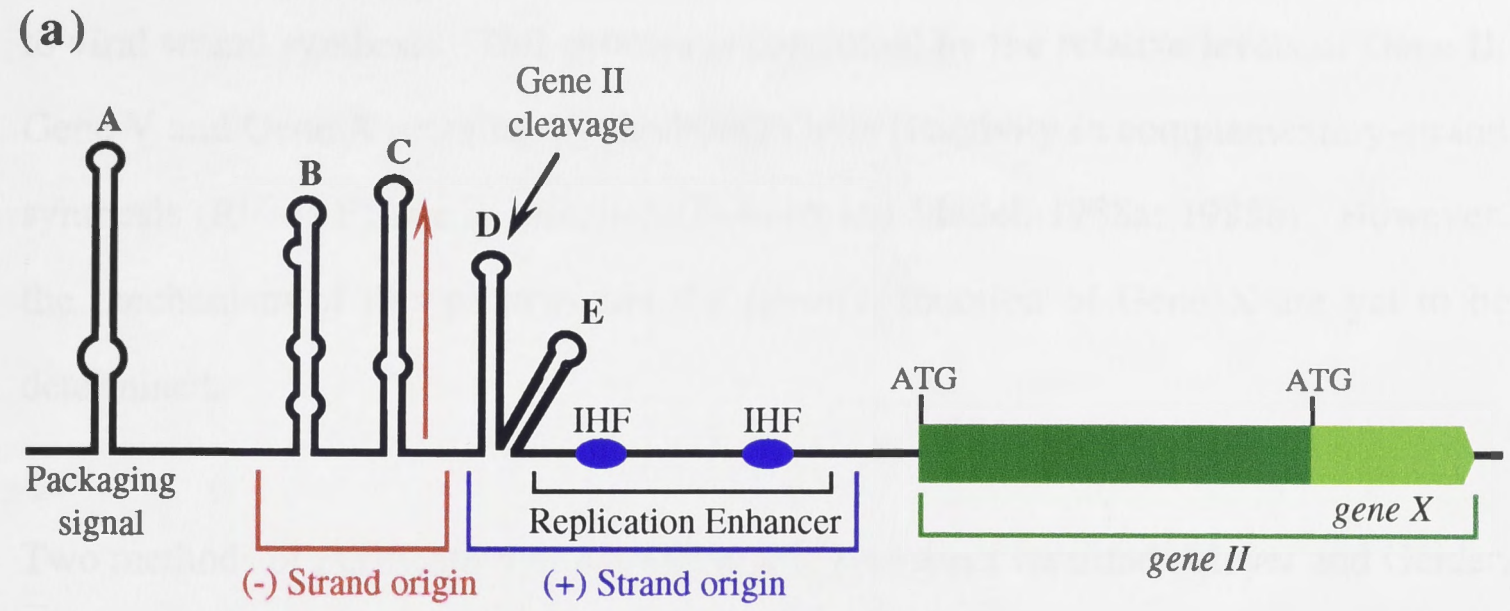
5.2 Introduction

Gene II protein is a 46-kDa endonuclease that is involved in initiation and termination of M13 DNA synthesis. It is the only replication protein encoded by the virus which relies on its host for the supply of the rest of the replication machinery. Gene II protein binds specifically to the (+) viral strand origin, presumably as a dimer, introducing a nick and providing a $3'\text{-}^{OH}$ DNA primer for rolling circle replication (Meyer et

al.,1979; Geider *et al.*, 1982) (Figure 5.1). Gene II is held tightly in a position that is favourable for a second cleavage and ligation of the nascent ssDNA once a round of replication has occurred. However, in contrast to the functionally analogous Gene A protein of the icosohedral phage ϕX174 (Chapter 6) no covalent linkage is formed with the 5´ end of the cleavage junction, nor is there any energy requirement for the ligation of the newly formed nascent single strand. It has been proposed that the energy from the second cleavage is in some way donated to the ligation (Geider *et al.*, 1982). Gene II is also known to interact with Rep helicase. This was established when RF molecules precleaved with Gene II could not be replicated, suggesting that a Gene II-Rep interaction is necessary for the formation of replication forks (Meyer and Geider, 1982).

In M13, *gene II* is located downstream of an intergenic region which contains both the (+) and (-) strand origins (van Wezenbeek *et al.*, 1980) (Figure 5.1). Five secondary structure hairpins (A-E) have been identified in this region (Zinder and Horiuchi, 1985). The cleavage site for Gene II is located within a hairpin structure (D) at position 5780 (between T and A of TTAA) of the (+) strand origin (Meyer *et al.*, 1979). The (+) strand origin also contains another hairpin (E), and a replication enhancer with binding sites for *E. coli* integration host factor (IHF) (Figure 5.1). High levels of Gene II can compensate for some mutation in the enhancer but not in the core structure, suggesting that Gene II interacts with the enhancer as well as its cleavage site (Figure 5.1). The (-) strand origin consists of two hairpins (B and C), one of which is the site (C) for RNA priming by the host RNA polymerase (Figure 5.1). Further upstream is another hairpin (A) which is the signal for phage packaging.

Gene X is a 12-kDa protein which is identical to the C-terminal 12 kDa of Gene II. This gene is transcribed from an internal, in-frame ATG start codon within *gene II* which is preceded by a potential RBS (van Wezenbeek *et al.*, 1980; Yen and Webster, 1981) (Figure 5.1). This protein is thought to be involved in the switch from RF→RF



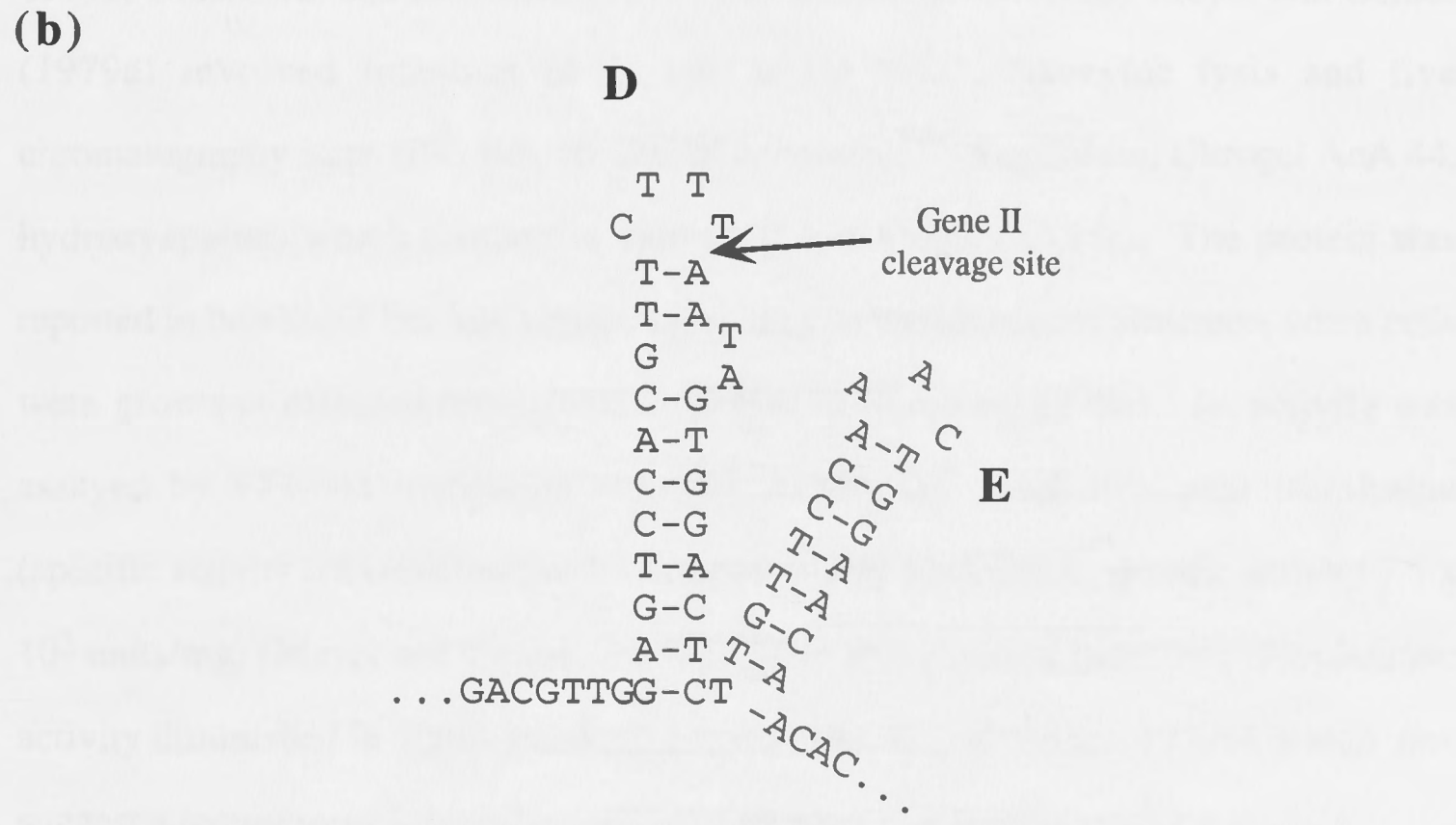


Figure 5.1

(a) A schematic representation of *gene II*, *gene X* and the intergenic region of M13. *Gene X* is transcribed from an internal in-frame ATG start codon within *gene II* (van Wezenbeek *et al.*, 1980). The intergenic region is located upstream of *gene II* and *gene X*, and contains the (+) and (-) strand origins. The (+) strand origin contains the cleavage site for Gene II protein (located in hairpin D) and two binding sites for *E. coli* IHF which are part of a replication enhancer (Greenstein *et al.*, 1988). The (-) strand origin consists of two hairpins (B and C) and RNA priming by RNA polymerase occurs at hairpin C (red arrow) (Zinder and Horiuchi, 1985).

(b) An enlargement of hairpins D and E showing the exact position of cleavage by Gene II protein between T and A bases at the top of hairpin D (Meyer *et al.*, 1979).

to viral strand synthesis. This process is controlled by the relative levels of Gene II, Gene V and Gene X proteins, by inhibiting Gene II activity in complementary-strand synthesis (RF→RF) late in infection (Fulford and Model, 1988a; 1988b). However, the mechanism of this process and the specific function of Gene X are yet to be determined.

Two methods of purification of native Gene II have been reported (Meyer and Geider, 1979a; Greenstein and Horiuchi, 1987). The method described by Meyer and Geider (1979a) involved infection of *E. coli* strain H402, lysozyme lysis and five chromatography steps (Bio-Rex 70, DEAE cellulose, DNA cellulose, Ultrogel AcA 44, hydroxyapatite) which resulted in extremely low yields (<0.1%). The protein was reported to be soluble but had a tendency to stick to membraneous structures when cells were grown at elevated temperatures (Meyer and Geider, 1979a). Its activity was assayed by RFI→ss replication (specific activity 0.7 μmol/mg), nick translation (specific activity 2.5 μmol/mg) and relaxation of RFI M13 DNA (specific activity 7.9 x 10^3 units/mg) (Meyer and Geider, 1979a). They also reported that Gene II replicative activity diminished in highly purified samples (Meyer and Geider, 1979a), which may suggest a requirement for another unidentified protein present in cruder materials.

Greenstein and Horiuchi (1987) improved the cellular levels of Gene II (30 mg/L of culture) following cloning of the gene into a runaway expression vector. Gene X was also overproduced to high levels. The Gene II protein was purified following lysis by sonication, treatment of the insoluble fraction with guanidine.HCl, and chromatography on Sephacryl S200 and S400 (to remove traces of Gene X). The yield of pure Gene II was <2% and the specific activities obtained were comparable with those reported by Meyer and Geider (1979a).

Due to the low yields obtained by the methods described above, a new method for purifying Gene II was necessary. The levels of its (and Gene X) overproduction has

been increased by cloning *gene II* into λ promoter vectors (developed in our group) and T7 promoter vectors (Studier *et al.*, 1990). Based on the observations reported by Meyer and Geider (1979a) and Greenstein and Horiuchi (1987), new methods of purifying Gene II have been developed. These methods are the focus of work presented in this chapter.

5.3 Materials and Methods

5.3.1 Cloning of gene II into pPL452

M13 *gene II* fortuitously has an *Nde*I restriction endonuclease site at its start codon. Wild-type M13 RF DNA was digested with *Bst*NI and the overhanging 5´ ends were made blunt with Klenow enzyme, then further digested with *Nde*I. A 1418-bp *Nde*I-*Bst*NI *gene II*$^+$ fragment (Meyer and Geider, 1979) was isolated. The plasmid pPL452 (Figure 2.1) was digested with *Nde*I and *Sma*I restriction enzymes and a 4210-bp fragment isolated. The two fragments were ligated at 16°C. Ampicillin-resistant transformants were selected at 30°C. Plasmid DNAs from transformants were isolated on a small scale and plasmids of the appropriate size (5628 bp) were digested with *Nde*I and *Kpn*I endonucleases to confirm the presence of 4208-bp and 1420-bp fragments. Overproduction of the desired proteins at 42°C was confirmed by SDS-PAGE. The constructed plasmid was designated pCL475 (Figure 5.2).

5.3.2 Construction of vectors for tagging proteins with an N-terminal His-tail

Immobilised metal chelate affinity chromatography has become a widely used technique for the purification of recombinant proteins (Schmitt *et al.*, 1993). The resin most commonly used is Ni(II)-NTA. The NTA (nitrilotriacetic acid) ligand linked to Sepharose CL-6B has 4 chelating sites which can interact with metal ions. Ni(II) ion

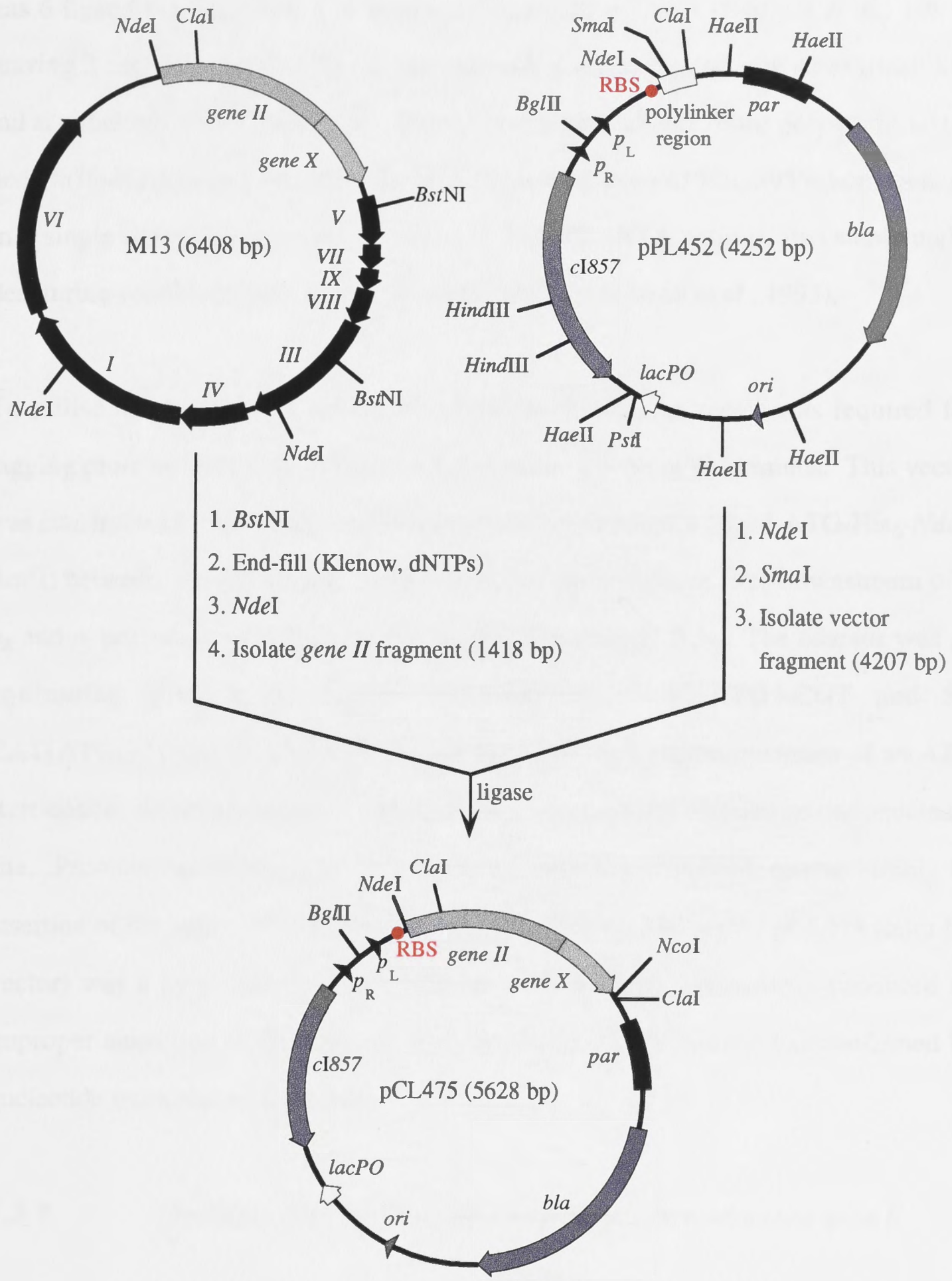


Figure 5.2

The scheme for cloning of M13 *gene II* into pPL452 for thermoinducible overproduction of its product from strong bacteriophage λ promoters. M13 was digested with *Bst*NI and ends were made blunt with the Klenow enzyme and dNTPs. After further digestion with *Nde*I, a 1418-bp *gene II*$^+$ fragment was isolated. This fragment was ligated with pPL452 previously digested with *Nde*I and *Sma*I to create pCL475.

has 6 ligand binding sites, 4 of which are occupied by NTA (Hochuli *et al.*, 1987), leaving 2 coordination sites free to interact with proteins tagged with a hexa-histidine tail at either the N- or C- terminus. The extremely high affinity of the poly-histidine tail and Ni(II)-NTA resin permits purification of proteins from <1% to >95% homogeneity in a single step (Janknecht *et al.*, 1991). The Ni(II)-NTA resin is also stable under denaturing conditions such as in 6 M guanidine.HCl (Schmitt *et al.*, 1993).

To utilise the Ni(II)-NTA resin for protein purification, a vector was required for tagging proteins with a hexa-histidine tail at either the N- or C-terminus. This vector was constructed by inserting a dsDNA oligonucleotide adaptor (*Hpa*I-ATG-His_6-*Nde*I-*Aat*II) between the *Hpa*I and *Aat*II restriction endonuclease sites downstream of λ p_R and p_L promoters of pPL451, creating pCL476 (Figure 5.3). The adaptor was an equimolar mixture of 5′-AATATTATG[CACCAT]$_3$ATGACGT and 5′-CAT[ATGGTG]$_3$CATAATATT containing an AT rich region upstream of an ATG start codon, followed by six histidine codons and an *Nde*I restriction endonuclease site. Proteins can be tagged with a hexa-histidine tail at the N-terminus simply by insertion of the gene at the unique *Nde*I site of pCL476. The vector pCL478 (tetra-his vector) was a by-product of the construction of pCL476, presumably produced by improper annealing of the adaptor. Construction of both vectors was confirmed by nucleotide sequence determination.

5.3.3 Cloning of gene II into pCL476 for construction of a His_6-gene II fusion

A fragment containing *gene II* (1418 bp) was excised from M13 DNA with restriction endonucleases *Nde*I and *Bst*NI as described in Section 5.3.1. The fragment was then ligated between the *Nde*I and *Sma*I sites of pCL476 to produce pCL477 (Figure 5.4). The ligation products were transformed into *E. coli* strain AN1459. Ampicillin-resistant transformants were selected and plasmid DNAs were isolated on a small scale.

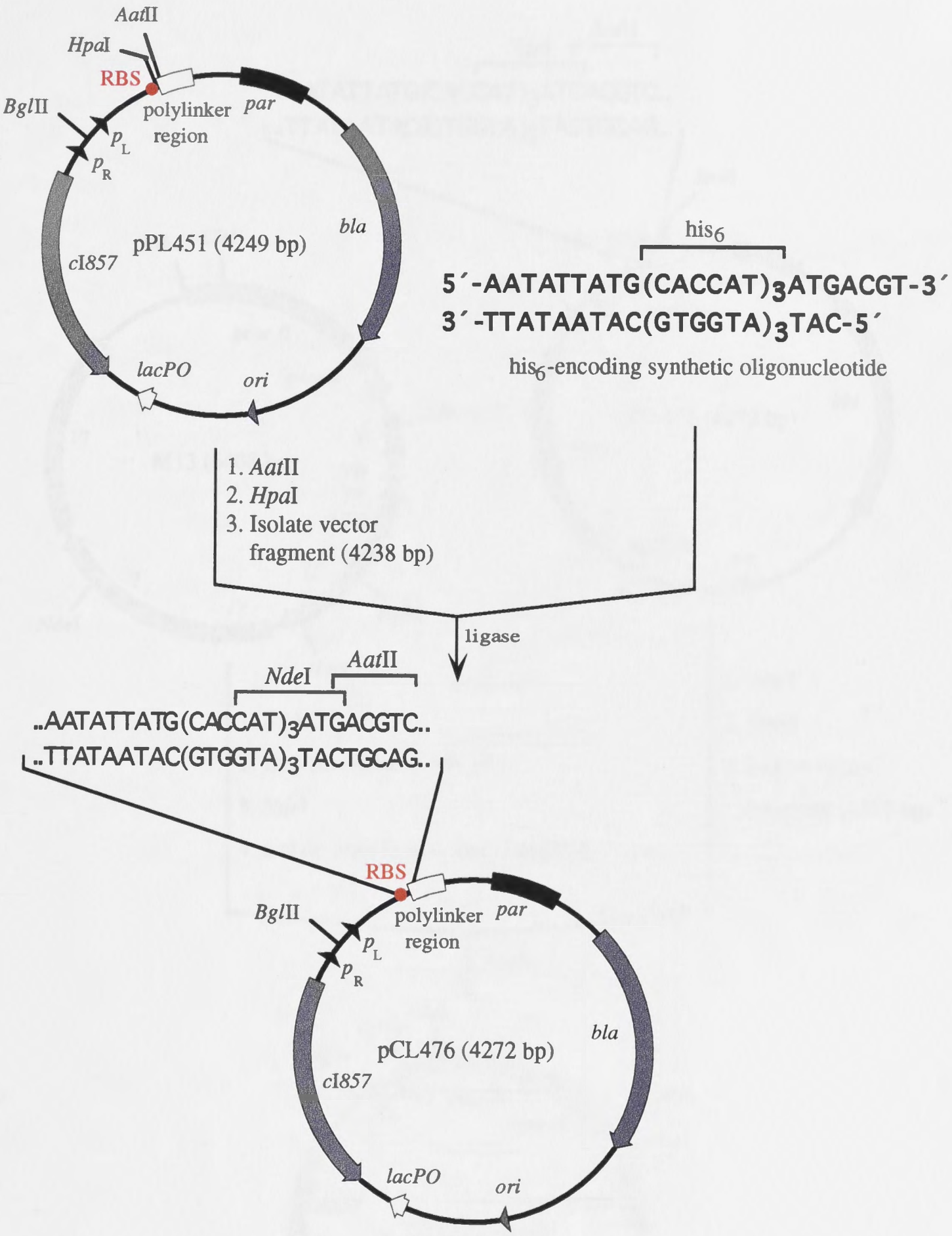


Figure 5.3

Construction of pCL476, a thermoinducible expression vector for cloning genes at a unique *Nde*I site which is preceded by a strong RBS, an ATG start codon and six histidine codons. Proteins overproduced are tagged at the N-terminus with a his_6 tail. The vector was constructed by ligating a ds-DNA oligonucleotide with an *Aat*II overhang at one end, between the *Aat*II and *Hpa*I sites of pPL451, resulting in construction of pCL476.

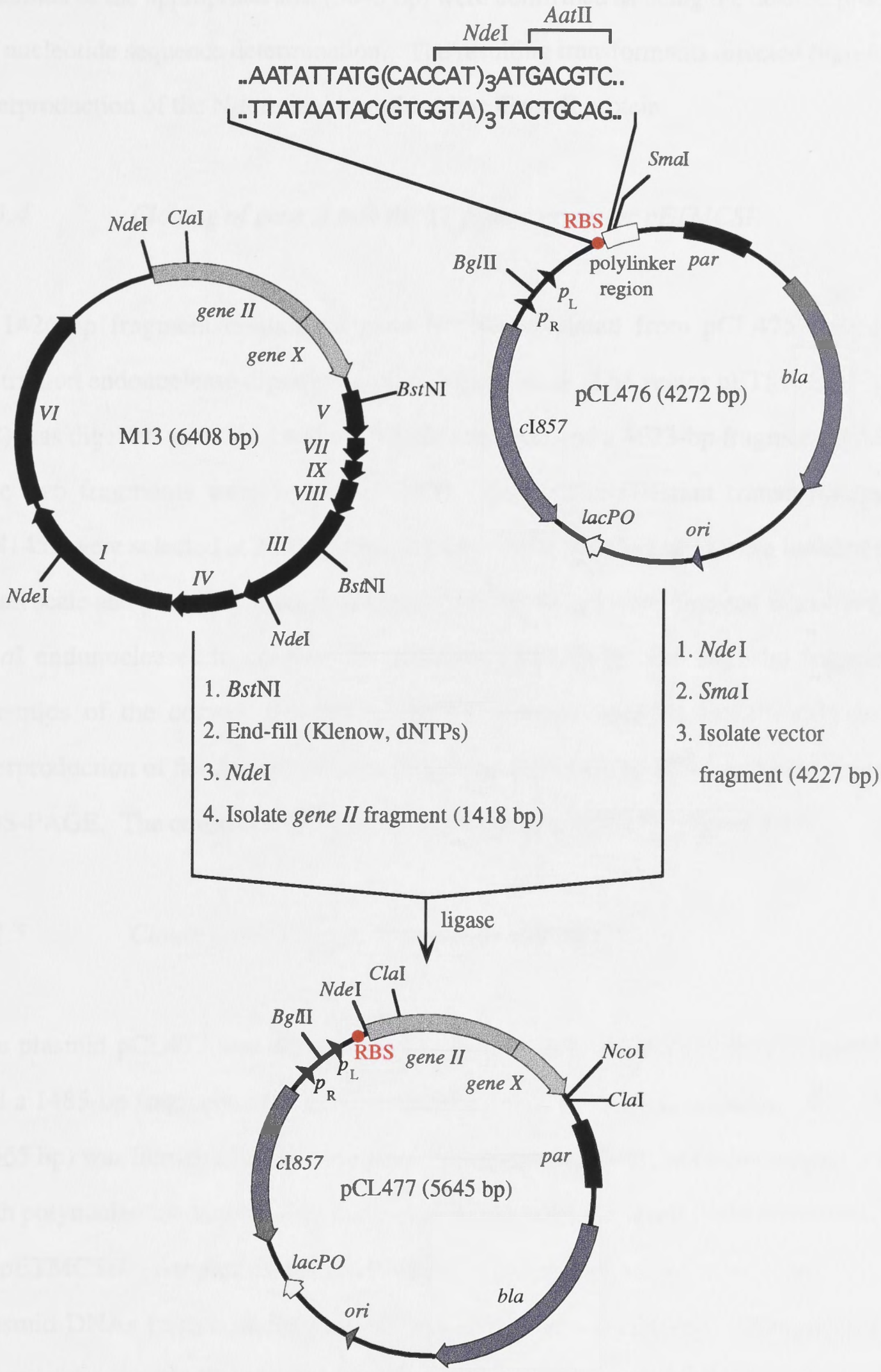


Figure 5.4

Construction of pCL477, a vector for thermoinducible overproduction of his_6-Gene II protein. The cloning was performed as described in Section 5.3.3.

Plasmids of the appropriate size (5645 bp) were confirmed as being the desired product by nucleotide sequence determination. The resulting transformants directed high-level overproduction of the N-terminal hexa-histidine Gene II protein.

5.3.4 Cloning of gene II into the T7 promoter vector pETMCSI

A 1426-bp fragment containing *gene II*$^+$ was isolated from pCL475 following restriction endonuclease digestion with *Nde*I and *Nco*I. The vector pETMCSI (Figure 2.2) was digested with *Nde*I and *Nco*I endonucleases and a 4623-bp fragment isolated. The two fragments were ligated at 16°C. Ampicillin-resistant transformants of AN1459 were selected at 37°C. Plasmid DNAs from transformants were isolated on a small scale and plasmids of the appropriate size (6049-bp) were digested with *Nde*I and *Nco*I endonucleases to confirm the presence of 4623-bp and 1426-bp fragments. Plasmids of the correct size were then transformed into BL21(DE3)/plysS and overproduction of the desired proteins following induction by IPTG was confirmed by SDS-PAGE. The constructed plasmid was designated pCL479 (Figure 5.5).

5.3.5 Cloning of His_6-gene II fusion into pETMCSI

The plasmid pCL477 was digested with restriction endonucleases *Bam*HI and *Bgl*II and a 1485-bp fragment containing the *His_6-gene II* fusion was isolated. pETMCSII (4665 bp) was linearised with restriction endonuclease *Bam*HI, then dephosphorylated with polynucleotide kinase. The *His_6-gene II* fragment was ligated into the *Bam*HI site of pETMCSII. Ampicillin-resistant AN1459 transformants were selected at 37°C. Plasmid DNAs from transformants were isolated on a small scale. Plasmids of the appropriate size (6150 bp) were digested with *Bam*HI and *Eco*RI endonucleases to confirm the presence of 4629-bp and 1521-bp fragments. These plasmids were then transformed into BL21(DE3)/plysS and overproduction of the desired proteins following induction by IPTG was confirmed by SDS-PAGE. The constructed plasmid

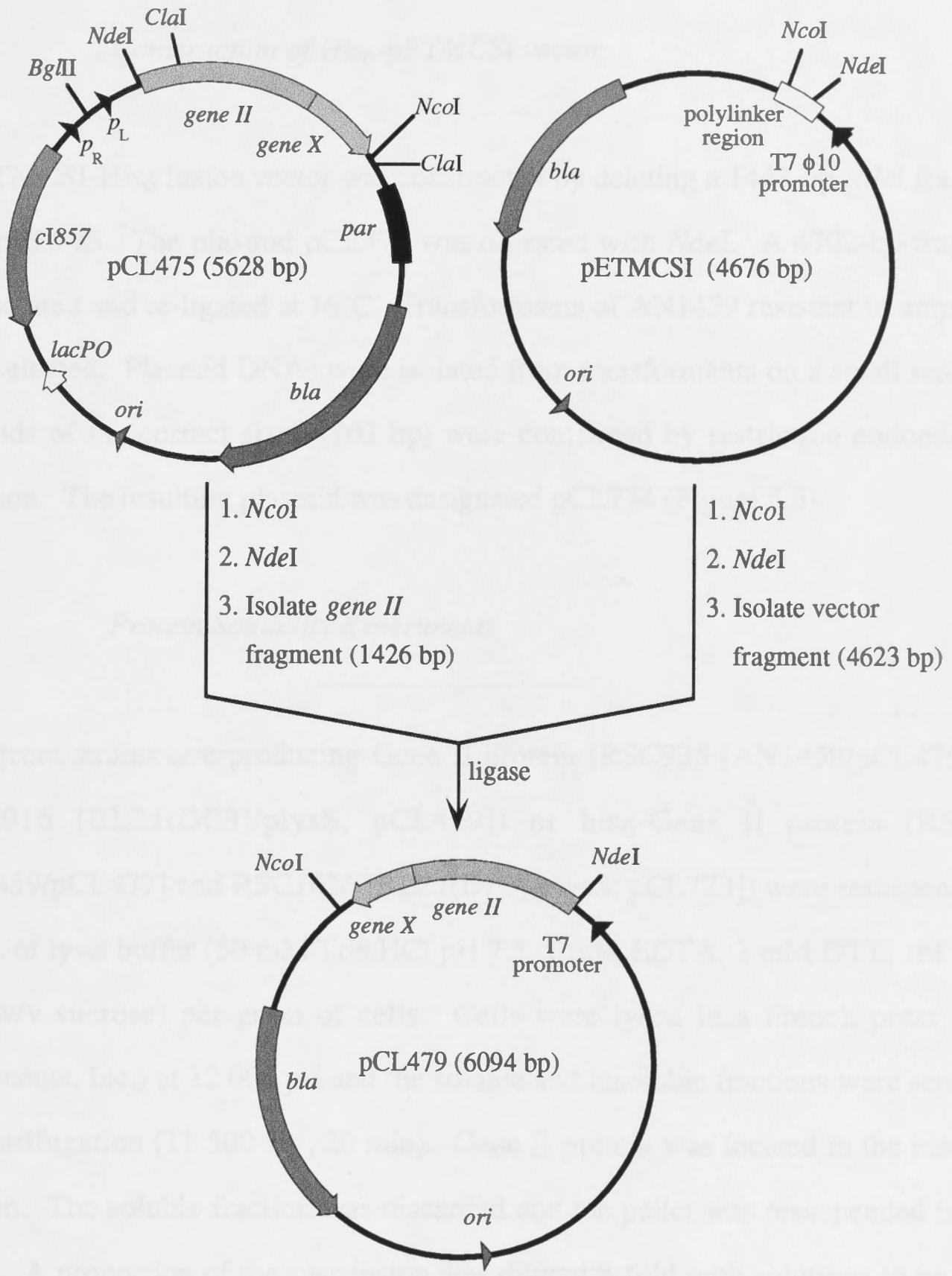


Figure 5.5

The construction of pCL479, a plasmid that directs IPTG inducible expression of *gene II* (and *gene X*) from a synthetic RBS upstream of an ATG start codon in *E. coli* strain BL21(DE3). A *gene II*$^+$ fragment was excised from pCL475 with restriction enzymes *Nde*I and *Nco*I and ligated between the corresponding sites of pETMCSI to create pCL479.

was designated pCL773 (Figure 5.6).

5.3.6 Construction of His$_6$-pETMCSI vector

A pETMCSI-His$_6$ fusion vector was constructed by deleting a 1447-bp *Nde*I fragment from pCL773. The plasmid pCL773 was digested with *Nde*I. A 4702-bp fragment was isolated and re-ligated at 16°C. Transformants of AN1459 resistant to ampicillin were selected. Plasmid DNAs were isolated from transformants on a small scale and plasmids of the correct size (4702 bp) were confirmed by restriction endonuclease digestion. The resulting plasmid was designated pCL774 (Figure 5.6).

5.3.7 Protein Solubility Experiments

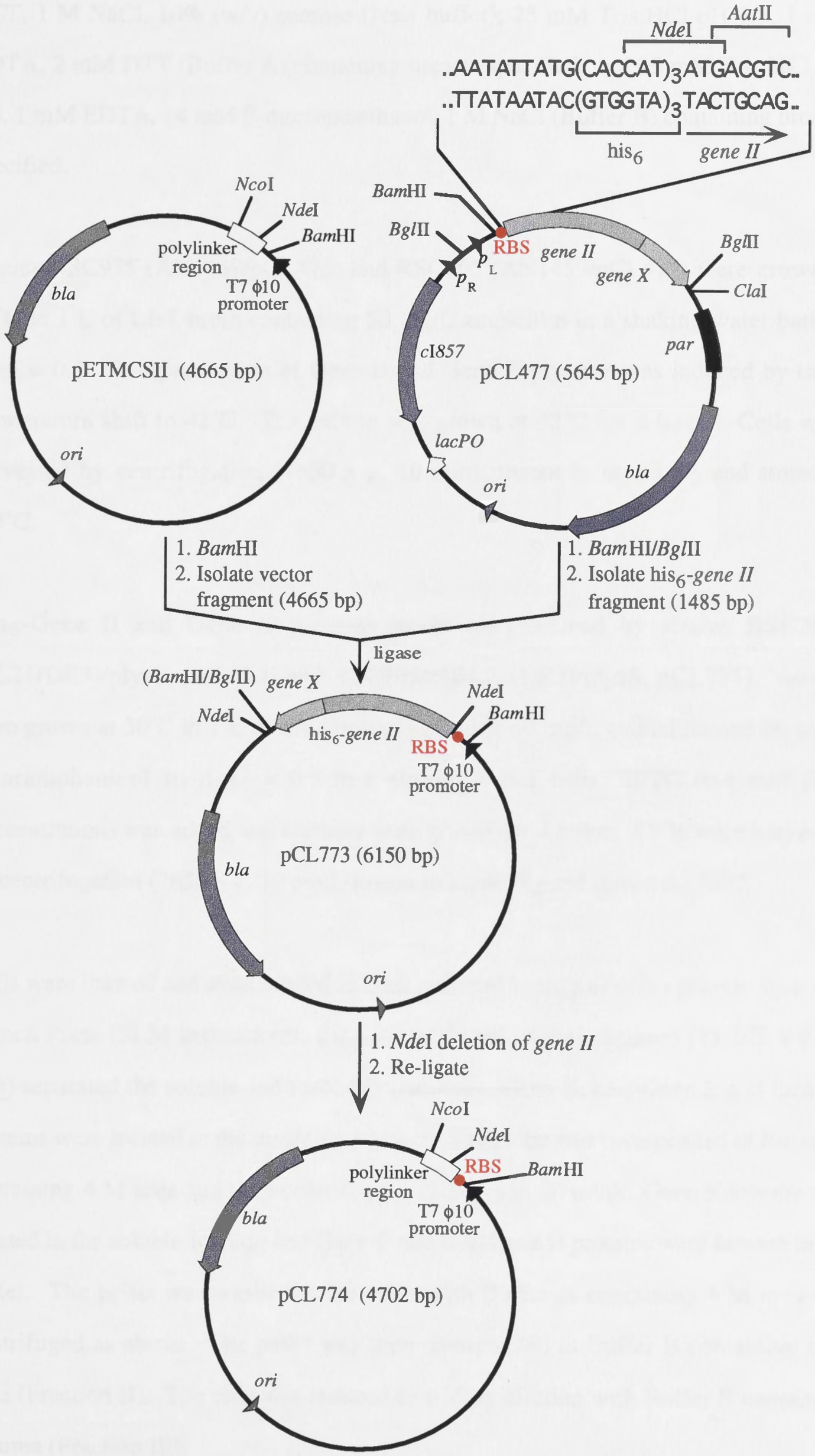
Cells from strains overproducing Gene II protein (RSC935 [AN1459/pCL475] and RSC2016 [BL21(DE3)/plysS, pCL479]) or his$_6$-Gene II protein (RSC965 [AN1459/pCL477] and RSC1056 [BL21(DE3)/plysS, pCL773]) were resuspended in 15 mL of lysis buffer (50 mM Tris.HCl pH 7.5, 1 mM EDTA, 2 mM DTT, 1M NaCl, 10% w/v sucrose) per gram of cells. Cells were lysed in a French press (SLM Instruments, Inc.) at 12 000 psi and the soluble and insoluble fractions were separated by centrifugation (11 500 x *g*, 20 min). Gene II protein was located in the insoluble fraction. The soluble fraction was discarded and the pellet was resuspended in lysis buffer. A proportion of the suspension was diluted 8-fold with solutions to be tested for solubilisation of Gene II protein. Suspensions were mixed and incubated at 4°C for 1 hour. The suspensions were clarified by centrifugation (microcentrifuge) and supernatants were analysed for soluble Gene II protein by SDS-PAGE.

5.3.8 Purification of Gene II or His$_6$-Gene II proteins

Buffers used in the purification were 25 mM Tris.HCl pH 7.5, 1 mM EDTA, 2 mM

Figure 5.6

The scheme for the construction of plasmid pCL773, which directs IPTG-inducible overproduction of his_6-Gene II (and Gene X), and pCL774, an IPTG-inducible expression vector for cloning genes at a unique *Nde*I site which is preceded by a strong RBS, an ATG start codon and a region that encodes for six histidine residues. pCL773 was constructed by excising a *gene II*$^+$ fragment from pCL477 with *Bam*HI and *Bgl*II and ligating it into the *Bam*HI site of pETMCSII. The vector pCL774 resulted from the removal of an *Nde*I fragment from pCL773 that contained *gene II* and recircularisation of the vector.


AatII
NdeI
..AATATTATG(CACCAT)3ATGACGTC..
..TTATAATAC(GTGGTA)3TACTGCAG..
his6
gene II
NcoI
NdeI
BamHI
polylinker region
T7 ϕ10 promoter
bla
ori
pETMCSII (4665 bp)
BamHI
BglII
RBS
gene II
pL
pR
gene X
BglII
ClaI
par
cI857
pCL477 (5645 bp)
lacPO
ori
bla
1. BamHI
2. Isolate vector fragment (4665 bp)
1. BamHI/BglII
2. Isolate his6-gene II fragment (1485 bp)
ligase
(BamHI/BglII)
gene X
NdeI
NdeI
BamHI
his6-gene II
RBS
T7 ϕ10 promoter
bla
ori
pCL773 (6150 bp)
1. NdeI deletion of gene II
2. Re-ligate
NcoI
NdeI
RBS
BamHI
polylinker region
T7 ϕ10 promoter
bla
ori
pCL774 (4702 bp)

DTT, 1 M NaCl, 10% (w/v) sucrose (lysis buffer); 25 mM Tris.HCl pH 7.5, 1 mM EDTA, 2 mM DTT (Buffer A) containing urea as specified; and 25 mM Tris.HCl pH 9.0, 1 mM EDTA, 14 mM β-mercaptoethanol, 1 M NaCl (Buffer B) containing urea as specified.

Strains RSC935 (AN1459/pCL475) and RSC965 (AN1459/pCL477) were grown at 30°C in 1 L of LBT broth containing 50 mg/L ampicillin in a shaking water bath to A_{595} = 0.5. Overproduction of Gene II and Gene X proteins was induced by rapid temperature shift to 42°C. The culture was grown at 42°C for 4 hours. Cells were harvested by centrifugation (7650 x *g*, 10 min), frozen in liquid N_2 and stored at -70°C.

His_6-Gene II and Gene X proteins were overproduced by strains RSC2016 [BL21(DE3)/plysS, pCL479] and RSC2056 [BL21(DE3)/plysS, pCL773]. Strains were grown at 30°C in 1 L of LBT broth containing 50 mg/L ampicillin and 30 mg/L chloramphenicol to A_{595} = 0.5 in a shaking water bath. IPTG (0.4 mM final concentration) was added and cultures were grown for 4 hours. Cells were harvested by centrifugation (7650 x *g*, 10 min), frozen in liquid N_2 and stored at -70°C.

Cells were thawed and resuspended in lysis buffer (15 mL/g of cells) prior to lysis in a French Press (SLM Instruments, Inc.) at 12 000 psi. Centrifugation (11 500 x *g*, 20 min) separated the soluble and insoluble fractions. Gene II, his_6-Gene II and Gene X proteins were located in the insoluble fraction. The pellet was resuspended in Buffer A containing 4 M urea and then centrifuged (11 500 x *g*, 20 min). Gene X protein was located in the soluble fraction and Gene II and his_6-Gene II proteins were located in the pellet. The pellet was washed twice more with Buffer A containing 4 M urea and centrifuged as above. The pellet was then resuspended in Buffer B containing 8 M urea (Fraction II). The urea was reduced to 1 M by dilution with Buffer B containing no urea (Fraction III).

5.3.9 *Purification of Gene II Protein by Metal Chelate Affinity Chromatography*

Recombinant his_6-Gene II protein to be purified by metal chelate affinity chromatography [Ni(II)-NTA] was prepared to the Fraction II stage as described in Section 5.3.9. Fraction II was dialysed against Tris.HCl pH 9.0, 5 mM β-mercaptoethanol containing 8 M urea (2 changes, 1 L/6 h) to remove traces of EDTA, then loaded onto a Ni(II)-NTA agarose column (QIAGEN) (15 x 1 cm), equilibrated with Tris.HCl pH 9.0 containing 8 M urea, at a flow rate of 0.3 mL/min. The urea concentration was reduced gradually from 8 to 1 M by a 150 mL gradient over 8.3 h. Bound protein was eluted with Tris.HCl pH 9.0, 1 M urea containing 0.25 M imidazole. Fractions containing his_6-Gene II were pooled (Fraction III).

5.3.10 *DNA Nicking Assay for Gene II Protein*

Endonuclease activity of Gene II protein was determined by nicking of covalently-closed supercoiled DNA containing the M13 or phage f1 origin. The DNAs used in the assay were M13 (Messing *et al.*, 1977) and pMA200 (Elvin *et al.*, 1990).

The assay was a modification of that described by Meyer and Geider (1979b) and Dotto *et al.* (1984). Reaction mixtures were prepared at 0°C and contained: 20 mM Tris.HCl pH 7.5; 100 μg/mL BSA; 8 mM DTT; 0.01% (w/v) Brij-58; 8 mM magnesium acetate; 125 mM potassium glutamate and either M13 (107 fmol) or pMA200 (107 fmol). Gene II protein to be enzymatically assayed was diluted in 50 mM Tris.HCl pH 7.5, 1 mM EDTA, 5 mM β-mercaptoethanol and either 6 M guanidine.HCl or 8 M urea. Aliquots (50 μL) of reaction mixture were prepared at 0°C. Upon addition of protein, the reactions were treated at 30°C for 10 min. Reactions were terminated by the addition of 12 μL of stop solution (0.16% bromophenol blue, 41.6% glycerol, 1.6% SDS). The reaction products were analysed by visualisation with ethidium bromide of

RFI (supercoiled) conversion to RFII (nicked) after electrophoresis through a 0.7% agarose gel in TAE buffer (Section 2.7).

5.3.11 RF→ss Replication Assay for Gene II Protein

Replication of M13 ssDNA from RFI was also used to determine the activity of Gene II protein. All proteins required for the assay were prepared in the group of Dr N. E. Dixon: the β subunit of DNA polymerase III holoenzyme (3 x 10^6 unit/mg) was prepared by Dr J. L. Beck; DNA polymerase III* (5.5 x 10^5 unit/mg), SSB (8.8 x 10^6 unit/mg), Rep helicase (1.1 x 10^7 units/mg) and M13 RFI DNA were obtained from Dr N. E. Dixon.

The assay was based on the *in vitro* replication of fd bacteriophage described by Meyer and Geider (1982). Reaction mixtures contained: 20 mM Tris.HCl pH 7.5; 100 μg/mL BSA; 8 mM DTT; 0.01% (w/v) Brij-58; 8 mM magnesium acetate; 125 mM potassium glutamate; 1 mM rATP; rCTP, rUTP, rGTP (250 μM each); 1 μg of SSB; Rep helicase; 100 ng of DNA polymerase III*; 26 ng of β subunit of DNA polymerase III holoenzyme; dATP, dCTP, dGTP, [^{3}H]dTTP (50 μM each) and M13 RFI (220 pmol, as nucleotide). The reaction mix was prepared at 0°C. Gene II protein was diluted in 50 mM Tris.HCl (pH 7.6), 0.1 mM EDTA, 2 mM DTT, 1 M urea. Aliquots (25 μL) of reaction mixture were prepared at 0°C and after addition of Gene II protein the reactions were treated at 30°C for 20 min. Cooling to 0°C and addition of 1 mL of stop solution (10% TCA, 0.1 M $NaPP_i$) terminated replication of DNA. After 10 min at 0°C, the mixtures were filtered through Whatman GF/C filters and washed with ~10 mL of 1 M HCl containing 0.1 M $NaPP_i$. Filters were then washed with ethanol, dried and counted in a Beckman LS 6000 IC scintillation counter. Ready-Safe scintillation cocktail was purchased from Beckman. Activity of Gene II protein in DNA replication was calculated in units (U), where one unit of replication activity denotes incorporation of one picomole of nucleotide into product per minute at 30°C.

5.4 Results and Discussion

5.4.1 Overproduction of Gene II Protein in λ and T7 Promoter Vectors

Strains containing plasmids pCL475 and pCL479 directed high-level overproduction of Gene II protein while strains containing plasmids pCL477 and pCL773 overproduced a his_6-Gene II fusion protein. Plasmids pCL475 and pCL477 are derivatives of pPL452 and pPL451, respectively. In both plasmids *gene II* has been placed under the control of bacteriophage λ promoters p_R and p_L and is preceded by a strong RBS derived from the vector. When strains RSC935 (AN1459/pCL475) and RSC965 (AN1459/pCL477) are grown at 30°C, the repressor is active and prevents transcription from the λ promoters. On shift of cultures to 42°C, the repressor is inactivated, and transcription from the promoters enabled (Figure 5.8).

The plasmids pCL479 and pCL773 were constructed from pETMCSI and pETMCSII, respectively, and *gene II* is under the control of the bacteriophage T7ϕ10 promoter. The strain BL21(DE3) contains the *lacI* gene, the *lac*UV5 promoter, the beginning of the *lacZ* gene and the T7 RNA polymerase gene on an insert in the chromosome on a stable lysogen of phage DE3. The *lac*UV5 promoter directs IPTG-inducible expression of the T7 RNA polymerase gene. T7 RNA polymerase then directs expression of plasmid-borne genes under the control of the T7 promoter. The plasmid plysS expresses low levels of T7 lysozyme, a natural inhibitor of T7 RNA polymerase which can be tolerated by *E. coli* because it is unable to pass through the inner membrane (Moffatt and Studier, 1987). Expression of T7 lysozyme eliminates the basal levels of T7 RNA polymerase to allow the cloning of genes that express products which are toxic to bacterial strains. Upon treatment of the cell with IPTG more T7 RNA polymerase is produced than can be inhibited by T7 lysozyme produced by plysS (Studier *et al.*, 1990). When these plasmids were transformed into strain BL21(DE3)/plysS the resulting strains RSC2016 [BL21(DE3)/plysS, pCL479] and

RSC2056 [BL21(DE3)/plysS, pCL773] directed high-level overproduction of Gene II and his_6-Gene II, respectively, from the T7 bacteriophage promoter upon addition of IPTG (Figure 5.8).

Gene II was overproduced following insertion of the gene into pPL452 under the control of the strong bacteriophage λ promoters p_R and p_L and downstream of a synthetic RBS which is perfectly complementary to the 3′ end of the *E. coli* 16S rRNA. The resulting plasmid (pCL475), when transformed into strain AN1459 (RSC935), directed high-level overproduction of the *gene II* product upon temperature shift from 30 to 42°C. Gene X protein, which is identical to the C-terminal 12 kDa of Gene II and translated from an internal ATG start codon was also overproduced to very high levels (Figure 5.7). The levels of Gene II and Gene X continued to increase up to five hours after temperature shift (Figure 5.7) without affecting growth of the cell culture.

The purification of overproduced proteins can be simplified to a one-step procedure by tagging proteins at either the N- or C-termini with a $(his)_6$ tail and using metal ion chelate affinity chromatography [eg. Ni(II)-NTA]. For this purpose, the vector pCL476 was constructed to tag proteins with six histidine residues at the N-terminus (Figure 5.3). This vector was then used for cloning of *gene II* (pCL477) for the overproduction of a his_6-tagged product. The strain RSC965 (AN1459/pCL477) overproduced similar levels of his_6-Gene II protein when compared to the overproduction of wild type Gene II (Figure 5.8).

Gene II (and Gene X) was overproduced on a large scale for a trial purification. The strain RSC935 (AN1459/pCL475) was grown at 30°C in 1 L of LB broth containing 50 mg/L ampicillin in a shaking water bath to A_{595} of 0.5. Overproduction of proteins was induced by rapid temperature shift to 42°C. The culture was grown at 42°C for 4

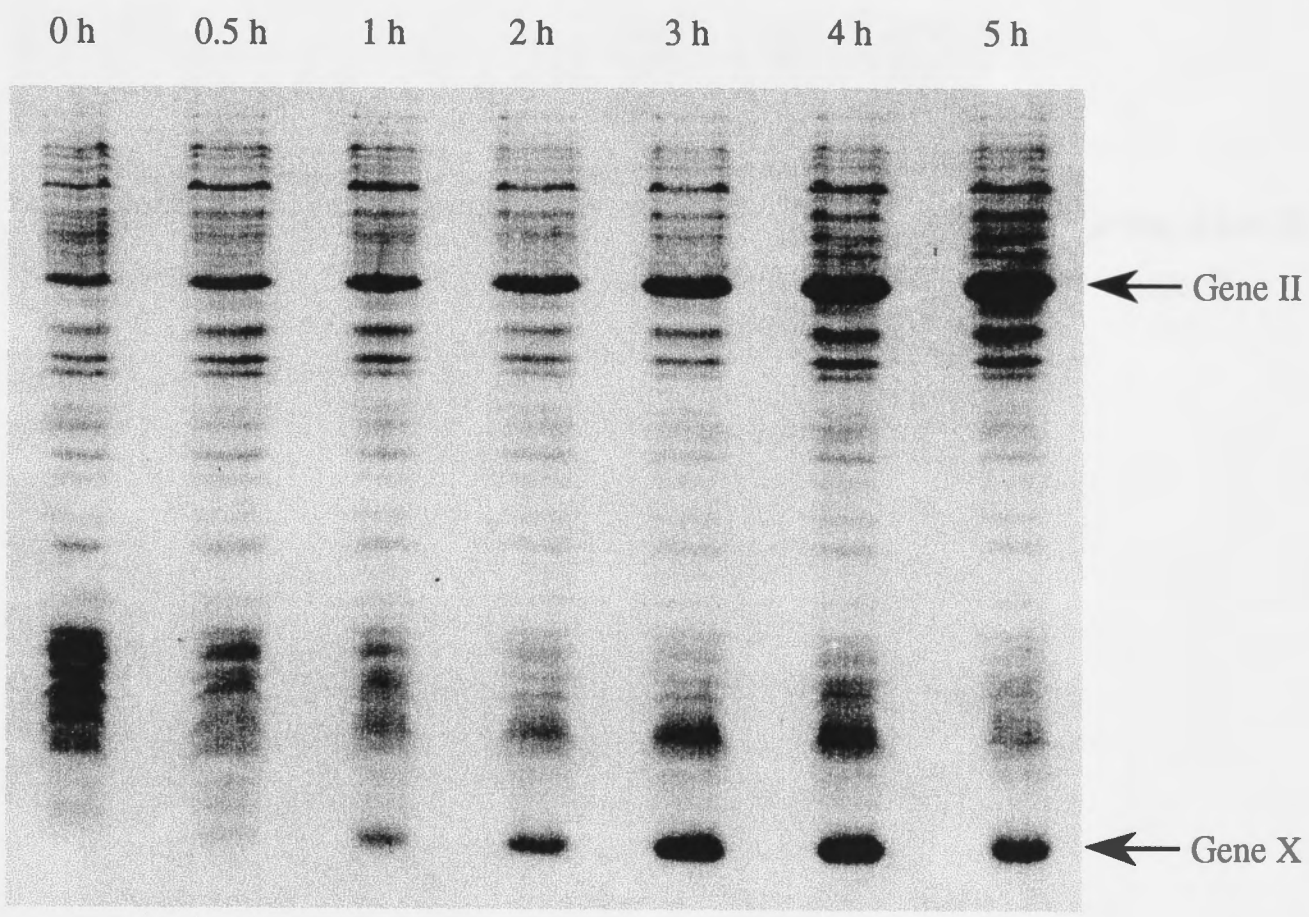


Figure 5.7

Overproduction of Gene II and Gene X proteins directed by strain RSC935 (AN1459/pCL475), detected by SDS-PAGE of lysed whole cells. Strain RSC935 was grown at 30°C in LB broth containing ampicillin to A_{595} = 0.5, then shifted to 42°C for 5 hours. Cells in 1 mL samples were harvested before (lane 0 h) and during treatment at 42°C (lanes 0.5, 1, 2, 3, 4 and 5 h), resuspended in an SDS loading buffer to an A_{595} of 10 and treated at 100°C (2 min) prior to application of 20 µL samples into lanes of a 15% SDS-PAGE slab gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue.

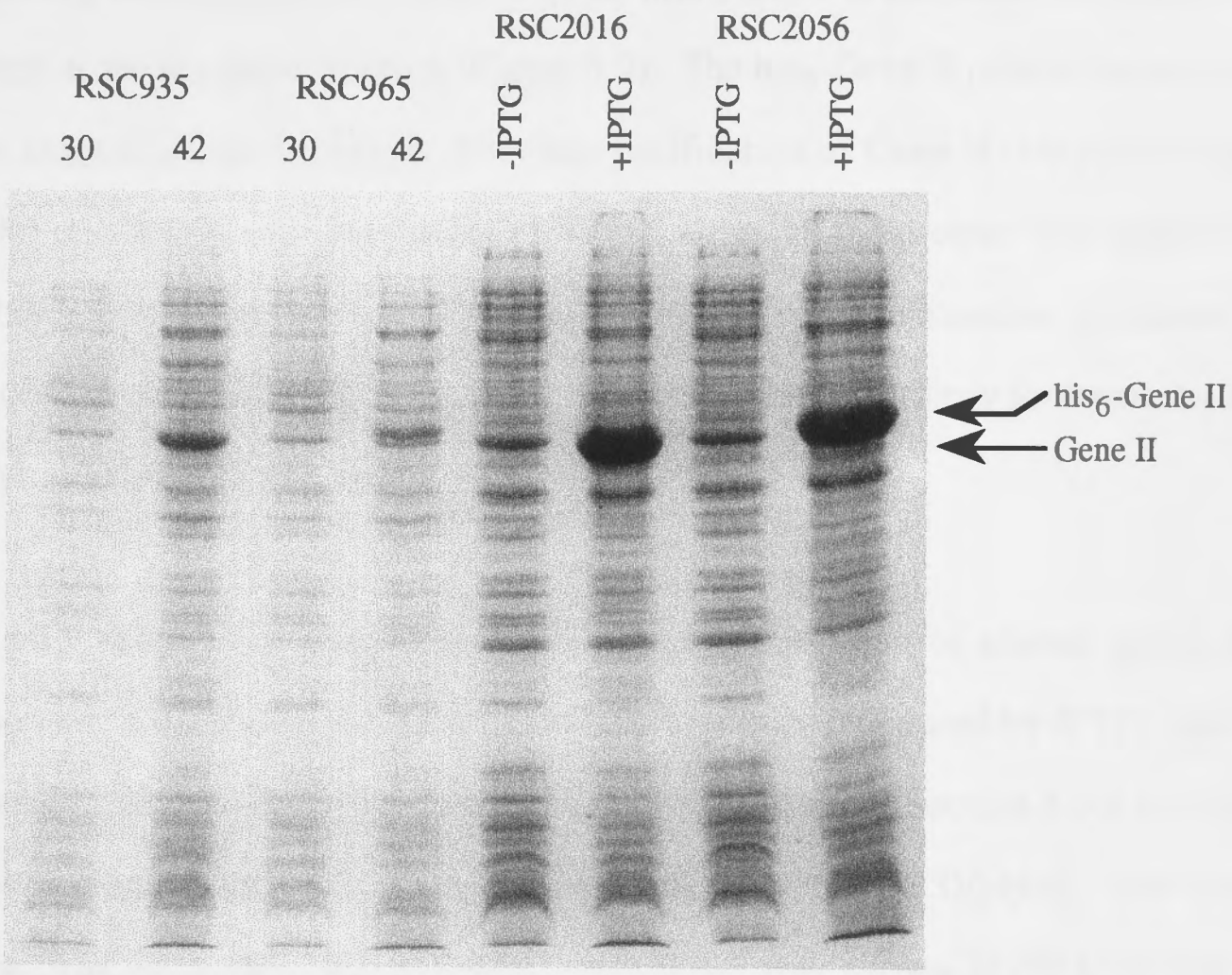


Figure 5.8

An SDS-PAGE showing the overproduction of Gene II and his_6-Gene II proteins from strains RSC935 (AN1459/pCL475), RSC965 (AN1459/pCL477), RSC2016 [BL21(DE3)/pCL479, pLysS] and RSC2056 [BL21(DE3)/pCL773, pLysS]. Strains RSC935 and RSC965 were grown at 30°C in LB broth containing ampicillin to A_{595} of 0.5, then shifted to 42°C and grown for 4 hours. RSC2016 and RSC2056 were grown at 30°C in LB broth containing ampicillin and chloramphenicol to A_{595} of 0.5, then IPTG was added and the cultures were grown for a further 4 hours. Cells in 1 mL samples were harvested before and after treatment, resuspended in SDS loading buffer to an A_{595} of 10, treated at 100°C (2 min) and 20 μL samples were loaded onto a 12% SDS-PAGE. Following electrophoresis, proteins were stained with Coomassie brilliant blue.

hours. Cells were harvested by centrifugation (7650 x *g*, 10 min), frozen in liquid N_2 and stored at -70°C. The cells were thawed and resuspended in lysis buffer (15 mL/g of cells) and lysed in a French Press (SLM Instruments, Inc.) at 12 000 psi. Following centrifugation (11 500 x *g*, 30 min), Gene II and Gene X proteins were located in the insoluble fraction (Figure 5.9). The his_6-Gene II protein behaved in the same manner as the wild type. Previous purification of Gene II (Meyer and Geider, 1979a) resulted in very low yield (>1%) and although the protein was reported to be soluble it was noted that it had a tendency to stick to membraneous structures when produced in cells grown at elevated temperatures. Thus, a strategy for low-temperature overproduction was required.

The pET vectors (Studier *et al.*, 1990) direct expression of cloned genes at low temperature from the bacteriophage T7 promoter which is induced by IPTG. M13 *gene II* was cloned into pETMCSI (Figure 2.2) as described in Section 5.3.4 to construct pCL479 (Figure 5.5) and transformed into strain BL21(DE3)/plysS. The resulting strain, RSC2016 [BL21(DE3)/plysS, pCL479], was grown at 30°C in LB broth containing ampicillin and chloramphenicol to an A_{595} of 0.5, then IPTG was added (to a final concentration of 0.4 mM) and the culture grown for a further four hours. Gene II (and Gene X) were overproduced to very high levels, >10 fold higher than that produced by RSC935 (based on SDS-PAGE analysis) (Figure 5.8). A pET vector for tagging cloned genes with an N-terminal his_6 tag was constructed (pCL774, Figure 5.6) and the levels of his_6-Gene II overproduction from strain RSC2056 [BL21(DE3)/plysS, pCL773] was as high as with the wild type (Figure 5.8).

Strain RSC2016 [BL21(DE3)/plysS, pCL479] was used for large-scale overproduction of Gene II protein for further purification trials. RSC2016 was grown at 30°C in 1 L of LB broth containing 50 mg/L ampicillin and 30 mg/L chloramphenicol in a shaking water bath to an A_{595} of 0.5. IPTG was added to a final concentration of 0.4 mM and the culture was grown for 4 hours. Cells were harvested by centrifugation (7650 x *g*,

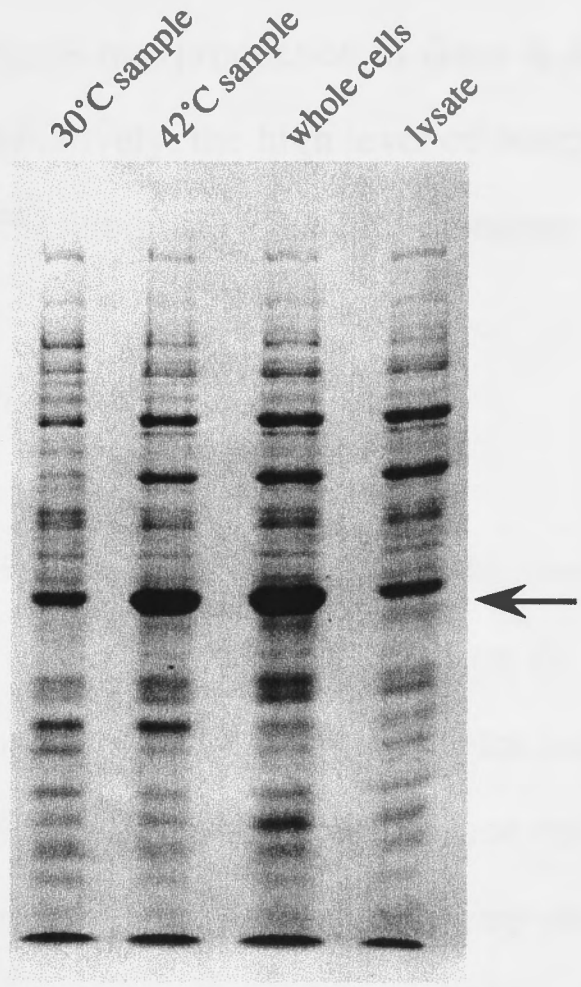


Figure 5.9

An SDS-PAGE of the overproduction of Gene II protein from strain RSC935 (AN1459/pCL475) and purification to the cell-free lysate stage. Strain RSC935 was grown at 30°C in LB broth containing ampicillin to A_{595} of 0.5, then shifted to 42°C and grown for 4 hours. Cells in 1 mL samples were harvested before and after treatment at 42°C, resuspended in SDS loading buffer to an A_{595} of 10, treated at 100°C (2 min) and 20 µL samples were loaded onto a 12% SDS-PAGE. SDS loading buffer was added to equivalent amounts of purification samples (whole cells and lysate), which were treated at 100°C (2 min) prior to loading on the same gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue. The arrow indicates the position of Gene II protein. The band at a similar position to Gene II in the 30°C sample and the lysate is an unidentified *E. coli* protein that co-migrates in the gel.

10 min), frozen in liquid N_2 and stored at -70°C. The cells were thawed and resuspended in lysis buffer (15 mL/g of cells) and lysed in a French Press at 12 000 psi. The soluble and insoluble fraction were separated by centrifugation (11 500 x *g*, 30 min). Once again Gene II and Gene X were located in the insoluble pellet (Figure 5.10). Thus, low temperature overproduction of Gene II did not assist in maintaining it in a soluble state. Alternatively, the high level of overproduction is unlikely to be favourable since Gene II insolubility is a result of monomer aggregation.

5.4.2 Solubilisation of Gene II Protein

Purification and charaterisation of Gene II require the protein to be in a soluble state. There are many factors which can have an effect on the solubility of proteins (eg. buffer, pH, etc.) and many techniques are available for rendering protein soluble (eg. chaperone folding, use of surfactants, etc.). A common method employed uses urea or guanidine.HCl to solubilise the protein followed by refolding on removal of the denaturants. Experiments were set up to determine the concentration of guanidine.HCl and urea required to solubilise Gene II. The protein was found to be fully soluble in either 8 M urea (Table 5.1) or 6 M guanidine.HCl (Table 5.2). The high concentration of denaturant required to solubilise Gene II is a reflection of its insolubility. However, Gene X protein was soluble at 4 M urea (Table 5.1) or 2 M guanidine.HCl (Table 5.2). Thus, Gene II and Gene X could be separated easily by resuspending the insoluble protein in 4 M urea or 2 M guanidine.HCl, then discarding the supernatant. After centrifugation, only Gene II was present in the insoluble material. This procedure also removes many other proteins that are soluble in low concentrations of urea and guanidine.HCl and may simplify the purification of Gene II.

A number of buffers at varying pH (4.7-9.5) were tested to determine conditions that would keep Gene II soluble following the removal of the denaturant(s). Experiments were carried out as described in Section 5.3.7. Gene II was found to be soluble in

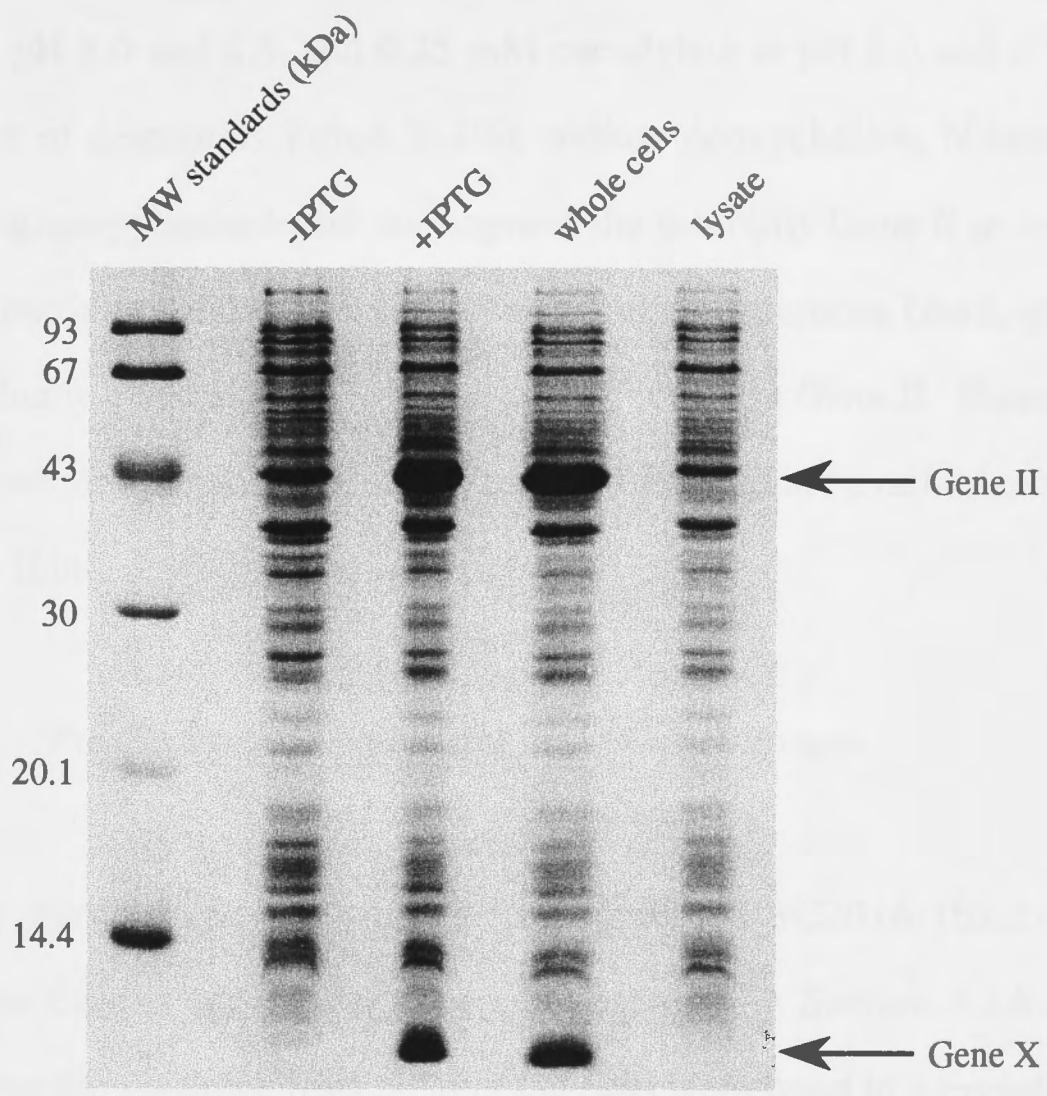


Figure 5.10

Overproduction of Gene II protein from strain RSC2016 [BL21(DE3)/plysS, pCL479] and purification to the cell-free lysate stage. Strain RSC2016 was grown at 30°C in LB containing ampicillin and chloramphenicol to an A_{595} of 0.5, then IPTG was added and the culture was grown for 4 h. Cells in 1 mL samples were harvested before (-IPTG) and after (+IPTG) induction, resuspended in SDS loading buffer to an A_{595} of 10, treated at 100°C (2 min) and 20 µL samples were loaded onto a 15% SDS-PAGE with equivalent amounts of samples of whole cells and lysate. Following electrophoresis, the proteins were stained with Coomassie brilliant blue.

0.25 mM Tris.HCl at pH 9.0 or 9.5 in the presence of 1 M urea (Table 5.3) and it was assumed that the protein is folded under these conditions. If all of the urea was removed, Gene II precipitated. A small proportion of Gene II was soluble in 0.25 mM Tris.HCl at pH 8.0 and 8.5, and 0.25 mM cacodylate at pH 6.0 and 6.5 (Table 5.3). The addition of detergents Triton X-100, sodium deoxycholate, N-lauroylsarcosine, and octyl-β-glucopyranoside did not improve the solubility Gene II in solutions where it was only partially soluble. The use of molecular chaperones DnaK and GroES-EL, used individually or combined, were not able to solublise Gene II. However, now that all five chaperones (GrpE, DnaJ, DnaK and GroES-EL) are available, experiments to refold Gene II into a soluble form should be carried out.

5.4.3 Purification of Gene II and his_6-Gene II Protein

Large-scale purification of Gene II from strains RSC2016 [BL21(DE3)/plysS, pCL479] was carried out by the procedure described in Section 5.3.8. Gene II was extracted from the insoluble fraction after the cells were lysed in a French press and the soluble and insoluble fractions separated by centrifugation. Gene X protein was located in the insoluble fraction. This pellet was washed several times with 4 M urea to remove Gene X and most of the other contaminating proteins. Gene II was then solubilised with Buffer B containing 8 M urea (Fraction II) and the urea concentration was reduced by diluting Fraction II 8-fold with Buffer B. This resulted in highly pure Gene II (>95% based on SDS-PAGE) (Figure 5.11 and Table 5.4).

His_6-Gene II was purified by the same method from strain RSC2056 [BL21(DE3)/plysS, pCL773] (Figure 5.11 and Table 5.4). The levels of Gene II and his_6-Gene II overproduced by strains RSC2016 and RSC2056 represented ~20-30% of the total protein of the cell. As a result 30-60 mg of protein could be obtained from 1 L of cell culture.

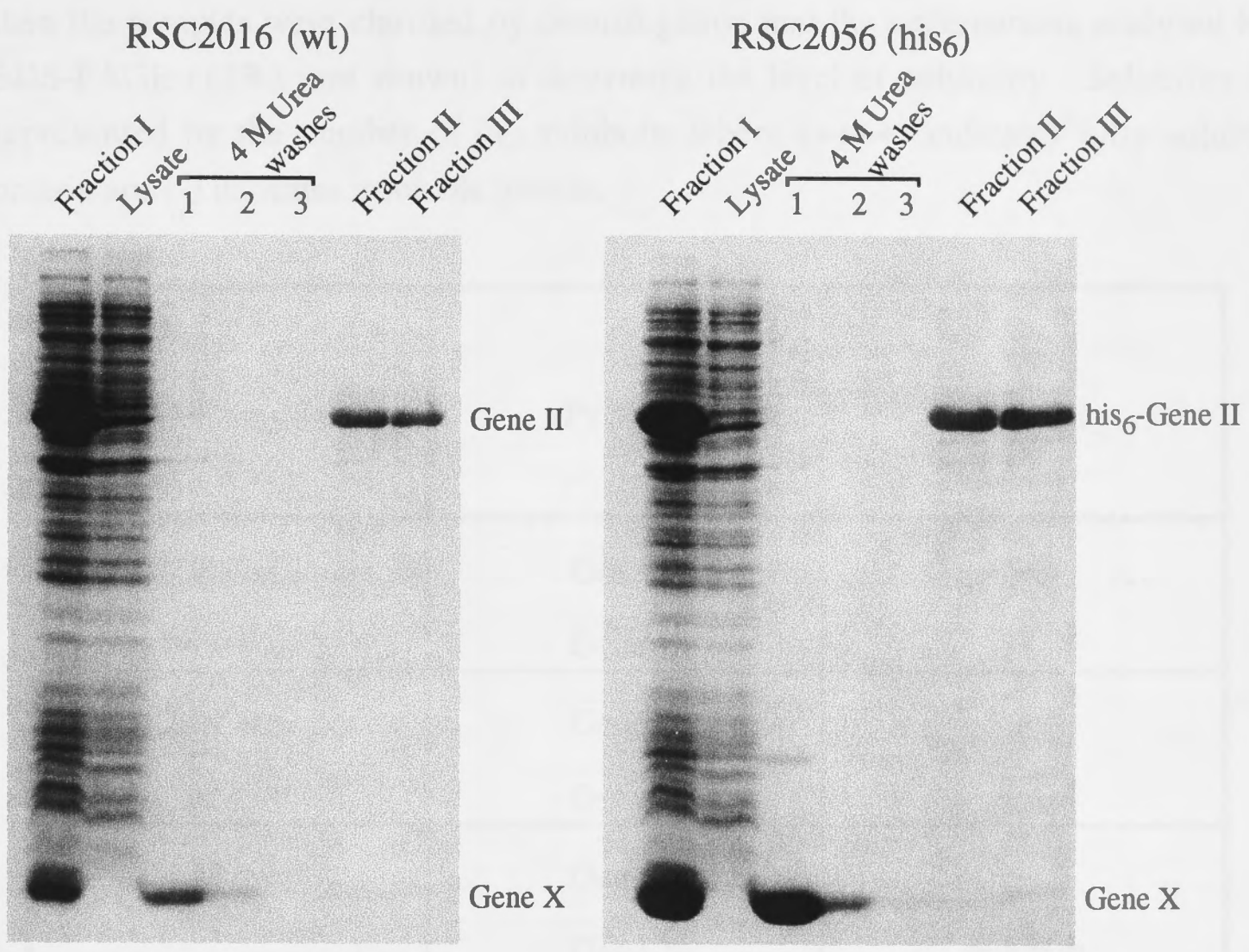


Figure 5.11

Purification of Gene II and his_6-Gene II proteins from strains RSC2016 [BL21(DE3)/plysS, pCL479] and RSC2056 [BL21(DE3)/plysS, pCL773]. The purification was carried out as described in Section 5.3.8. Proteins in Fractions and washes were separated by SDS-PAGE (15%). Equivalent amounts of each Fraction were loaded onto the gel. Following electrophoresis, proteins were stained with Coomassie brilliant blue.

Table 5.1

Solubility of Gene II and Gene X in buffers containing 25 mM Tris.HCl pH 7.5, 1 mM EDTA and 14 mM β-mercaptoethanol plus the components listed. The insoluble pellet after cell lysis was resuspended in the buffers indicated and treated at 0°C for 1 h, then the samples were clarified by centrifugation and the supernatants analysed by SDS-PAGE (15%) (not shown) to determine the level of solubility. Solubility is represented by the number of (+) symbols, where (++++) indicates fully soluble protein and (-) indicates insoluble protein.

Buffer Components	Protein	Solubility
2 M NaCl	Gene II	-
	Gene X	-
3 M urea	Gene II	-
	Gene X	++
4 M urea	Gene II	-
	Gene X	++++
5 M urea	Gene II	-
	Gene X	++++
6 M urea	Gene II	+
	Gene X	++++
7 M urea	Gene II	++
	Gene X	++++
8 M urea	Gene II	++++
	Gene X	++++

Table 5.2

Solubility of Gene II and Gene X in buffers containing 25 mM Tris.HCl pH 7.5, 1 mM EDTA and 14 mM β-mercaptoethanol plus the components listed. The insoluble pellet after cell lysis was resuspended in the buffers indicated and treated at 0°C for 1 h, then the samples were clarified by centrifugation and the supernatants analysed by SDS-PAGE (15%) (not shown) to determine the level of solubility. Solubility is represented by the number of (+) symbols, where (++++) indicates fully soluble protein and (-) indicates insoluble protein.

Buffer component	Protein	Solubility
1 M guanidine.HCl	Gene II	-
	Gene X	+++
2 M guanidine.HCl	Gene II	-
	Gene X	++++
3 M guanidine.HCl	Gene II	+
	Gene X	++++
4 M guanidine.HCl	Gene II	++
	Gene X	++++
5 M guanidine.HCl	Gene II	+++
	Gene X	++++
6 M guanidine.HCl	Gene II	++++
	Gene X	++++

Table 5.3

The effect of buffer and pH on the solubility of Gene II protein. The insoluble pellet after cell lysis was resuspended in an 8 M urea solution containing 1 mM EDTA and 2 mM DTT and diluted (8-fold) with the buffers indicated, then treated at 0°C for 1 h. The samples were clarified by centrifugation and the supernatants analysed by SDS-PAGE (15%) (not shown) to determine solubility. Solubility is represented by the number of (+) symbols, where (++++) indicates fully soluble protein and (-) indicates insoluble protein.

Buffer	pH	Components	Solubility
25 mM Na acetate	4.7	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM Na citrate	5.0	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM Na citrate	5.5	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM Na citrate	6.0	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM MES	6.0	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM MES	6.5	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM cacodylate	6.0	1 mM EDTA & 14 mM β-mercaptoethanol	+
25 mM cacodylate	6.5	1 mM EDTA & 14 mM β-mercaptoethanol	+
25 mM MOPS	6.8	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM Tris.HCl	7.5	1 mM EDTA & 14 mM β-mercaptoethanol	-
25 mM Tris.HCl	8.0	1 mM EDTA & 14 mM β-mercaptoethanol	+
25 mM Tris.HCl	8.5	1 mM EDTA & 14 mM β-mercaptoethanol	++
25 mM Tris.HCl	9.0	1 mM EDTA & 14 mM β-mercaptoethanol	+++
25 mM Tris.HCl	9.5	1 mM EDTA & 14 mM β-mercaptoethanol	+++

5.4.4 *Refolding and Purification of His_6-Gene II using Ni(II)-NTA Chromatography*

Insoluble proteins that have been solubilised using denaturants can often be refolded. This is usually achieved by carrying out the refolding in very dilute solutions, in large volumes, to minimise monomer interactions. However, this can complicate the recovery of protein. The monomer interactions could be minimised by immobilising the proteins on a solid support. The solid support would be required to bind the protein reversibly but not inhibit its folding or recovery. Proteins have been able to be refolded on solid supports such as agarose beads (Sinha and Light, 1975), ion-exchange matrices (Creighton, 1986) and metal chelate affinity matrices (Itoh *et al.*, 1996; Tuan *et al.*, 1996). The protein can be bound under denaturing conditions and folded by slowly altering the solvent to remove the denaturant. Ni(II)-NTA is a metal chelate affinity matrix which can bind proteins that are tagged with a hexa-histidine tail. Such a matrix is ideal for immobilised refolding of proteins as the polypeptide chain is held at one end and the support is unlikely to interfere with folding. The disadvantage with ion exchange matrices is that the polypeptide may be bound at more than one position and the residues that are making the contacts with the matrix need to be free for the protein to fold.

Cloning of *gene II* into a vector for tagging its product with an N-terminal his_6-tail was carried out as described in Sections 5.3.3 and 5.3.5. The his_6-Gene II product was overproduced at very high levels (Figure 5.8) and purified using Ni(II)-NTA chromatography as described in Section 5.3.9. Insoluble his_6-Gene II was solubilised in an 8 M urea solution containing 25 mM Tris.HCl pH 9.0, 5 mM β-mercaptoethanol, then loaded onto Ni(II)-NTA column pre-equilibrated with 25 mM Tris.HCl pH 9.0, 5 mM β-mercaptoethanol, 8 M urea. About half of the protein bound to the column. The urea concentration was reduced to 1 M over a 150 mL gradient (8.3 h) to allow the protein to fold slowly. The protein was eluted with Tris.HCl pH 9.0, 5 mM

Table 5.4

Purification of Gene II and his_6-Gene II by the method described in Section 5.3.8. The protein concentration was determined using the method of Bradford (Section 2.14). Cells were derived from 1 L of cell culture (2 and 2.5 g, wet weight, respectively).

	Gene II	his_6-Gene II
Fraction	**Protein** (mg)	**Protein** (mg)
I Whole Cells	171.3	217.2
II Urea Solubilisation	42.9	73.6
III Urea Dilution	28.6	63.6

β-mercaptoethanol, 1 M urea containing 0.25 M imidazole (Figure 3.12). Based on SDS-PAGE analysis, the protein was judged to be >95% pure.

5.4.4 Nicking Activity of Gene II Protein

The specific activity of Gene II can be measured in a nicking reaction that converts M13 RFI (supercoiled) to RFII (nicked or relaxed). Gene II purified here was assayed for nicking activity using the procedure described in Section 5.3.10. The nicking assay was followed over time to determine if Gene II was reused in the reaction. An assay over 90 min using Gene II (0.2 μg, ~60 molecules/M13 circle) showed that the cleavage increased and the reaction proceeded almost to completion (Figure 5.13). Similar results were obtained for his_6-Gene II purified using the same method. However, the process appears to be inefficient when considering the number of Gene II molecules used in the reaction and Gene II was only active when diluted into the assay from Fraction II of the purification (with 8 M urea present). Fraction(s) III Gene II and his_6-Gene II purified by Ni(II)-NTA chromatography were unable to cleave M13. It is possible that when the urea was diluted to 1 M that Gene II folded incorrectly. An alternative explanation is that when Gene II is diluted in the presence of its DNA target, some molecules are able to use the DNA as a scaffold for folding into the active conformation and these molecules remain active.

Based on the specific activity of Gene II nicking obtained by Meyer and Geider (1979a), approximately 2.8 molecules of Gene II were required per M13 molecule for the complete conversion of RFI to RFII (Meyer and Geider, 1979a). Nicking of M13 by Gene II purified here occurred at levels above 5 molecules of Gene II per molecule of M13 (Figure 5.14). However, full conversion of RFI to RFII did not occur even when Gene II was present at ~90 molecules/M13 (Figure 5.14). Some other factor (e.g. IHF) may be required that would increase the efficiency of M13 cleavage. This theory is yet to be tested.

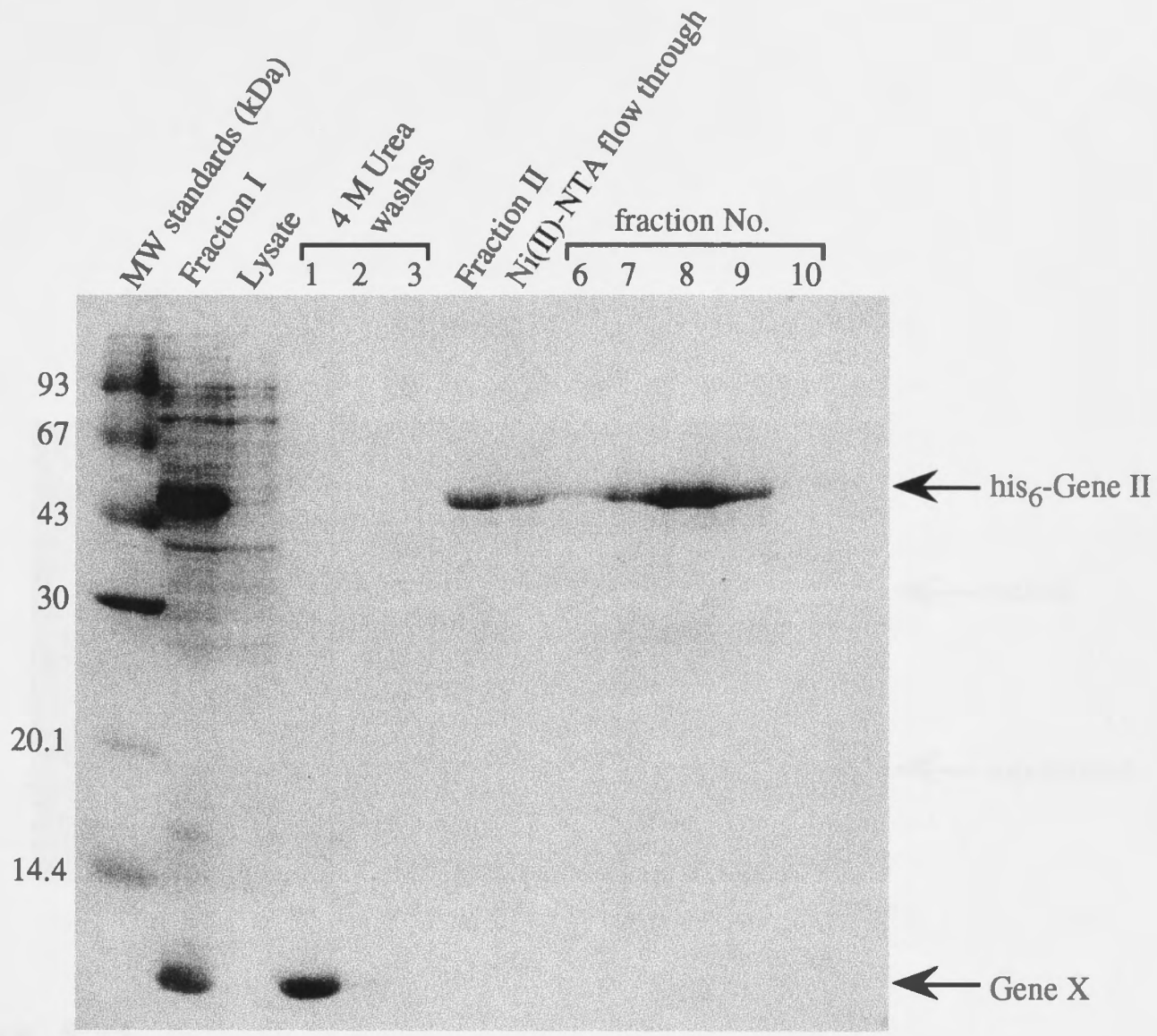


Figure 5.12

Purification of his_6-Gene II protein from strain RSC2056 [BL21(DE3)/plysS, pCL773) using Ni(II)-NTA metal chelate chromatography. The purification was carried out as described in Section 5.3.9. Proteins in purification Fractions, washes and Ni(II)-NTA chromatography fractions were separated by SDS-PAGE (15%). Following electrophoresis proteins were stained with Coomassie brilliant blue.

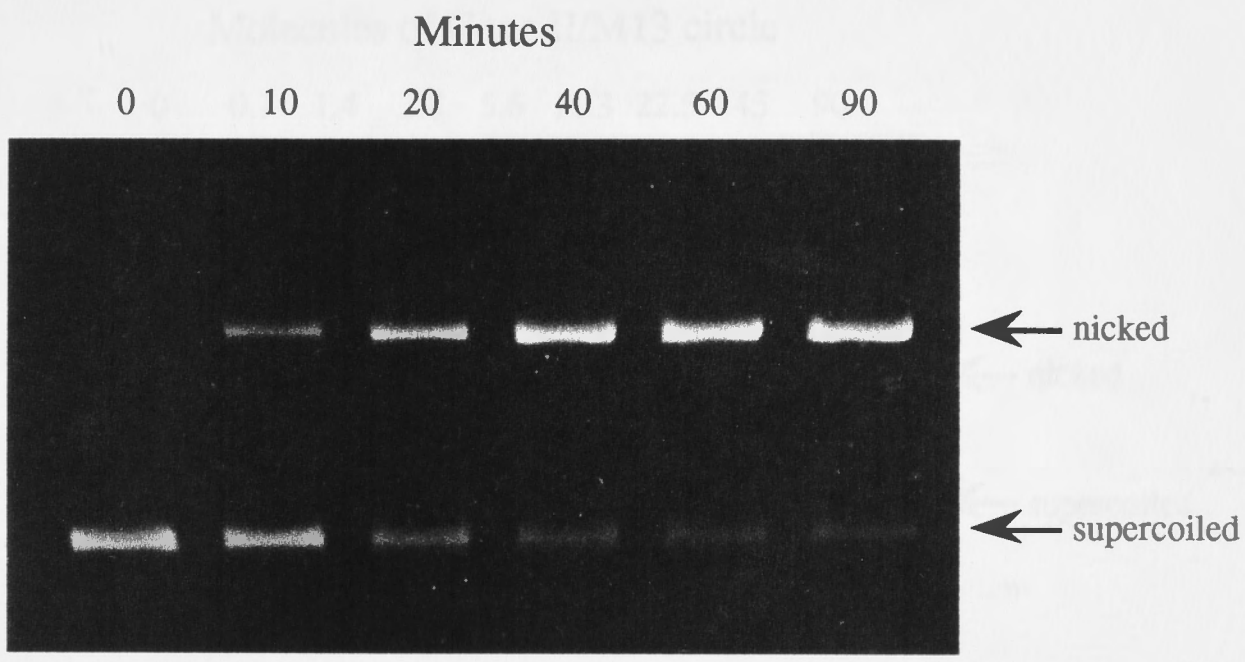


Figure 5.13

A 0.7% agarose gel of M13 RFI DNA nicked by Gene II protein over 90 min. Gene II protein (0.2 μg) was added to a reaction mix containing 20 mM Tris.HCl pH 7.5, 100 μg/mL BSA, 0.01% (w/v) Brij 58, 8 mM Mg-acetate, 125 mM K-glutamate and M13 supercoiled DNA (107 fmol). The reaction was placed at 30°C and samples were collected at 0, 10, 20, 40, 60 and 90 min. Stop mix (0.16% bromophenol blue, 41.6% glycerol, 1.6 % SDS) was added to the samples which were loaded onto the agarose gel. Following electrophoresis the DNA was stained with ethidium bromide (2 μg/mL).

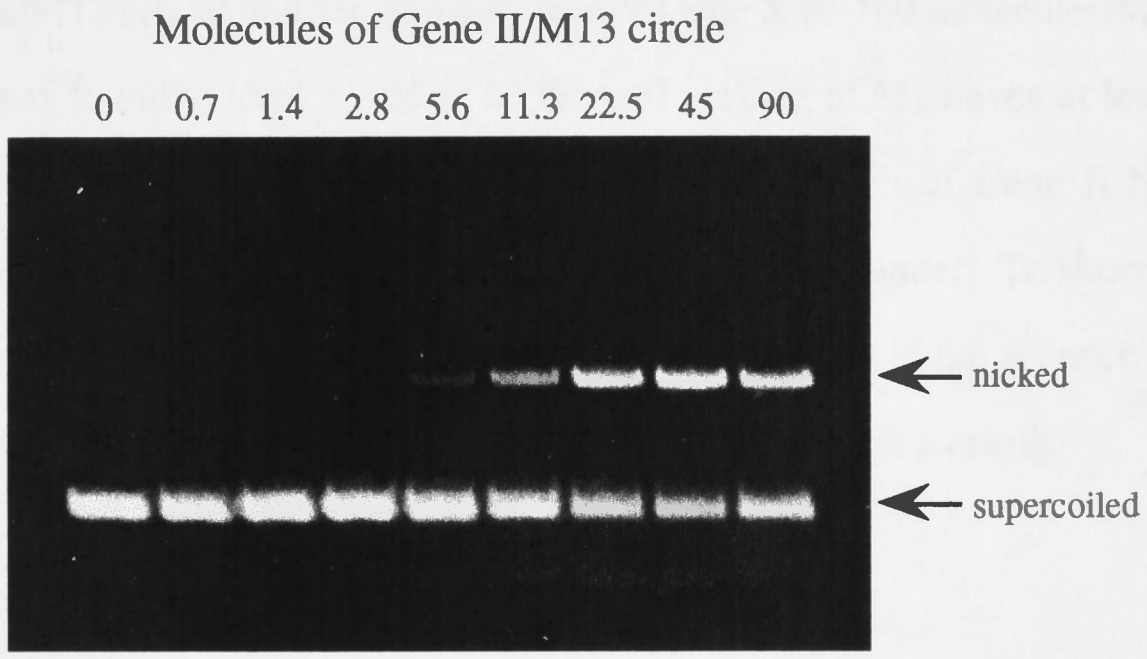


Figure 5.14

A 0.7% agarose gel of M13 nicked by Gene II protein. Gene II protein (0-90 molecules/circle) was added to a reaction mix containing 20 mM Tris.HCl pH 7.5, 100 µg/mL BSA, 0.01% (w/v) Brij 58, 8 mM Mg-acetate, 125 mM K-glutamate and M13 supercoiled DNA (107 fmol). The reaction mixtures were placed at 30°C for 10 min, then reactions were terminated by addition of stop mix (0.016% bromophenol blue, 41.6% glycerol, 1.6 % SDS). The samples were loaded onto the agarose gel and, following electrophoresis, the DNA was stained with ethidium bromide (2 µg/mL).

5.5.5 *The effect of Gene X protein on Gene II nicking of M13 RFI*

Gene X protein has been reported to be required for viral strand synthesis by inhibiting Gene II activity in complementary strand synthesis (Fulford and Model, 1984; 1988a). Experiments were performed to determine whether Gene X had an effect on Gene II cleavage activity. The nicking reaction described in Section 5.3.10 was used and reactions were performed with a constant amount of Gene II (67 ng, ~20 molecules/M13 circle) and varying amounts of Gene X (0-160 molecules/M13 circle). Gene X was found to have no effect on Gene II nicking of M13 even at levels of 160 molecules/M13 circle (Figure 5.15). Therefore, inhibition of Gene II by Gene X during M13 replication *in vivo* must occur in some other manner. To show that Gene X was not cleaving M13 during the reaction it was assayed in the absence of Gene II and was found to be unable to cleave M13 RFI DNA (data not shown).

5.5.6 *RF→ss Replication Activity of Gene II Protein*

The assay used here to measure the activity of Gene II in DNA synthesis of ss from RFI was described in Section 5.3.11 and was a modification of that described by Meyer and Geider (1982). Very little Gene II activity was seen in this assay. Several factors may explain the lack of activity. Possibly, Gene II purified here is not folded correctly and, thus, is unable to perform its replicative function. However, it was shown that Gene II did possess nicking activity which would suggest that some of the molecules are folded in the correct conformation.

Alternatively, another factor may be required for the RS→ss replication. Meyer and Geider (1979a) reported that very little Gene II activity was seen in the complementation assay for RF replication when highly purified protein was used. This suggests that another factor required for M13 RF replication has yet to be identified and is removed during purification. This may also explain the inefficiency of

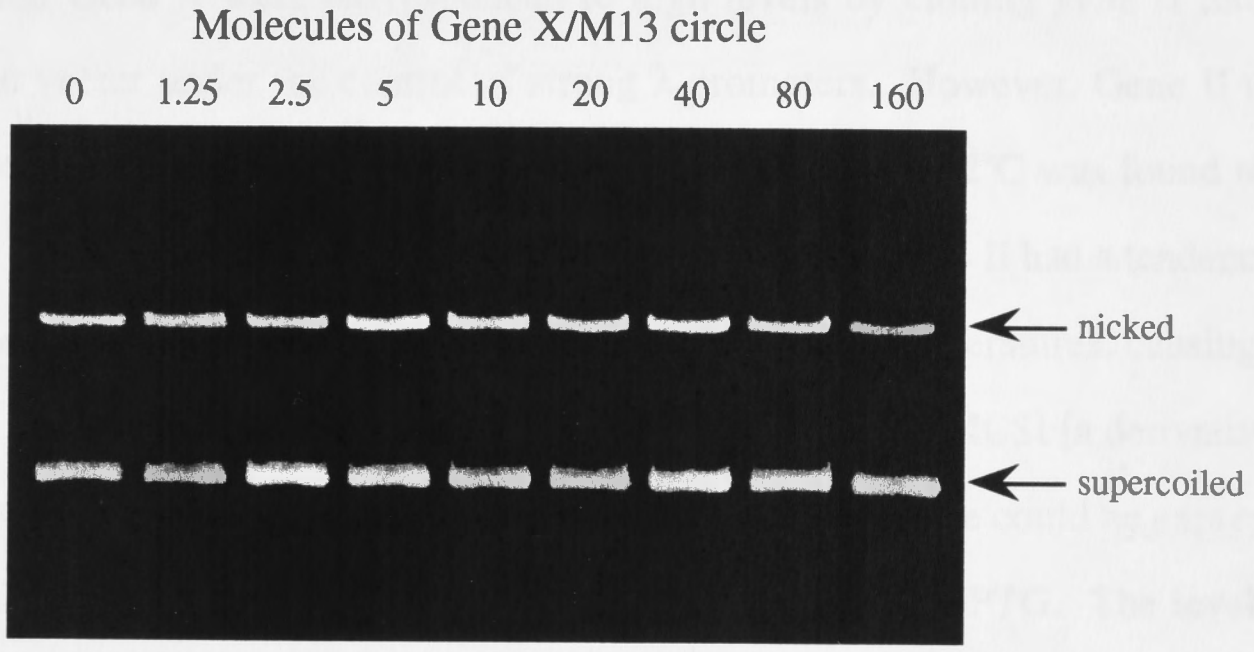


Figure 5.15

A 0.7% agarose gel of M13 nicking by Gene II in the presence of varying amounts of Gene X protein. Gene X protein (0-160 molecules/circle) was added to a reaction mixture containing 20 mM Tris.HCl pH 7.5, 100 µg/mL BSA, 0.01% (w/v) Brij 58, 8 mM Mg-acetate, 125 mM K-glutamate, M13 supercoiled DNA (107 fmol) and Gene II protein (20 molecules/circle). The reaction mixtures were placed at 30°C for 10 min, then reactions were terminated by the addition of stop mix (0.16% bromophenol blue, 41.6% glycerol, 1.6% SDS). The samples were loaded onto the agarose gel and, following electrophoresis, the DNA was stained with ethidium bromide.

Gene II in the nicking of M13. The (+) strand origin contains a replication enhancer region which contains binding sites for *E. coli* IHF. It is plausible to speculate that IHF may play a role in RF replication of M13.

5.5 Conclusions

Gene II and Gene X were overproduced to high levels by cloning *gene II* into an expression vector under the control of strong λ promoters. However, Gene II (and Gene X) overproduced in strains containing these plasmids at 42°C was found to be insoluble. It was reported by Meyer and Geider (1979a) that Gene II had a tendency to stick to membraneous structure when expressed at elevated temperatures, causing the protein to be insoluble. As a result, *gene II* was cloned into pETMCSI [a derivative of the pET vectors described by Studier et al. (1990)] where the gene could be expressed at low temperature from the T7ϕ10 promoter upon addition of IPTG. The levels of *gene II* expression were increased 10-fold compared to the expression in the λ promoter vectors and although Gene II represented 20-30% of the total cell protein, it remained insoluble. Vectors were constructed to tag proteins at the N-terminus to simplify their purification using metal chelate affinity chromatography. The level of his_6-Gene II overproduction was similar to that observed with the native protein, but the protein was still insoluble.

Extensive screening was undertaken to determine conditions to render Gene II and his_6-Gene II soluble. It was found that to fully solubilise the proteins 8 M urea or 6 M guanidine.HCl were required. Gene X protein, which was also overproduced, was soluble in 4 M urea or 2 M guanidine.HCl. The denatured protein was diluted to ascertain whether certain buffers, pH and anion surfactants could maintain Gene II in a soluble state. Gene II and his_6-Gene II were found to be fully soluble in 25 mM Tris.HCl at pH 9.0-9.5 containing 1 mM EDTA, 14 mM β-mercaptoethanol and 1M urea. It was assumed that the protein would be folded in 1 M urea.

A method for purifying Gene II on a large scale was developed which utilised the solubility of Gene II, his_6-Gene II and Gene X in different concentrations of urea. Insoluble, overproduced protein was washed several times with 4 M urea to remove Gene X and most of the other contaminating proteins. The insoluble Gene II or his_6-Gene II was solubilised with 8 M urea, then diluted into 25 mM Tris.HCl at pH 9.0, 1 mM EDTA, 14 mM β-mercaptoethanol. Approximately 30-60 mg of protein was obtained from 1 L of cell culture and was > 95% pure as judged by SDS-PAGE.

Gene II or his_6-Gene II was found to be active in relaxation of M13 RFI DNA at greater than 5 molecules of protein/circle of M13 but the activity was only observed when Gene II or his_6-Gene II was diluted into the assay directly from solutions containing 8 M urea. It is plausible to speculate that some Gene II and his_6-Gene II molecules are able to fold into their active conformation in the presence of their DNA target and these molecules remain active. Also, complete relaxation of M13 RFI DNA was never observed even when Gene II or his_6-Gene II was present at levels greater than 90 molecules/ M13 circle. Neither of these proteins were active in RF replication of M13. Meyer and Geider (1979a) reported that in highly purified samples, Gene II replicative activity diminished. Thus, it was suspected that an unknown factor removed during purification is required for RF replication. This could also explain the inefficiency of Gene II in the nicking of M13 RFI DNA.

Attempts were made to refold his_6-Gene II from a denatured state into its active conformation while immobilised on a Ni(II)-NTA resin. His_6-Gene II was bound to the column under denaturing conditions, then the urea reduced from 8 to 1 M using a gradient over 8.3 h. Soluble protein was eluted from the column. However, this protein was also inactive in relaxation of M13 RFI DNA. Further attempts to refold Gene II and his_6-Gene II using molecular chaperones (GroES-EL, DnaK, DnaJ and GrpE) are being carried out.

CHAPTER 6

CLONING OF THE ϕX174 gene A AND gene A*

6.1 Aims and Significance

The Gene A protein of bacteriophage ϕX174 is a multi-functional replication endonuclease, involved in the initiation and termination of the RF→ss stage of ϕX174 DNA replication (Eisenberg *et al.*, 1976; 1977). It functions by cleavage and ligation of replication intermediates without the expenditure of energy (Henry and Knipper, 1974; Eisenberg *et al.*, 1976; van Mansfeld *et al.*, 1978). Gene A also forms a covalent link with the 5′-^{P} end of the DNA and interacts with Rep helicase to form a Gene A-DNA-Rep complex during replication (Ikeda *et al.*, 1979, Eisenberg and Kornberg, 1979; Scott *et al.*, 1977; Kornberg *et al.*, 1978; Arai and Kornberg, 1981e). Another gene (*gene A**) is encoded within *gene A* from an internal in-frame start and the product (Gene A*) is involved in shutoff of host DNA replication and the switch from RF→RF to RF→ss replication of ϕX174 (Linney and Hayashi, 1973; Martin and Godson, 1975; Funk and Snover, 1976; Langeveld *et al.*, 1979, 1981; Eisenberg and Ascarelli, 1981).

There has been no reported cloning of *gene A*. Van der Avoort *et al.* (1983) reported that they were unable to clone *gene A* and speculated that expression of *gene A** may be lethal as a result of arresting cell DNA synthesis. The aim of work presented in this chapter was to investigate why *gene A* has been unable to be cloned and design a strategy for cloning of the gene.

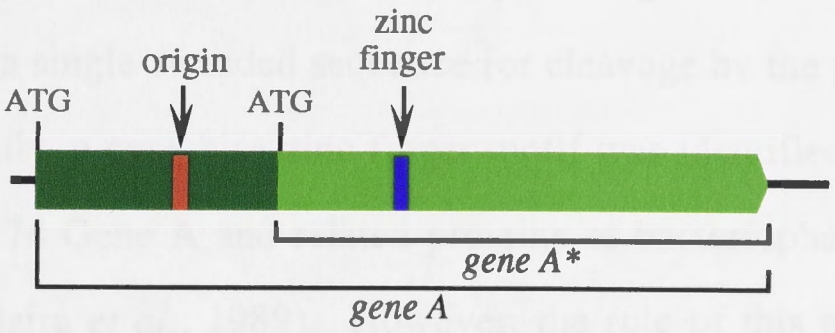
6.2 Introduction

ϕX174 Gene A protein is a 60-kDa replicative endonuclease which functions in a similar manner to Gene II protein of M13 (see Section 5.2). It is the most extensively studied replicative endonuclease. Like Gene II, Gene A is involved in the cleavage and ligation of replication intermediates and is the only replication protein encoded by the phage, which relies on its host for the supply of the other replication machinery (see

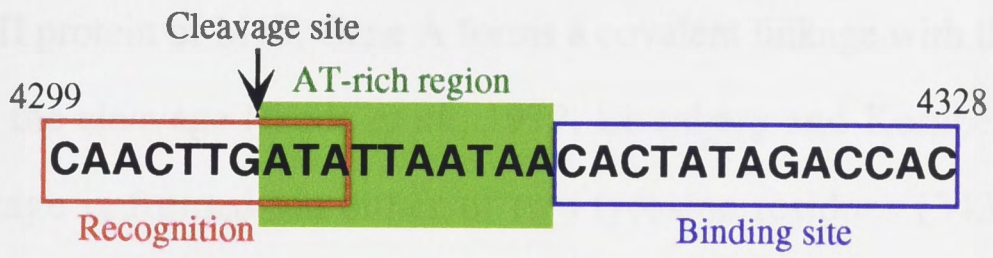
Section 5.2) (Henry and Knipper, 1974; Eisenberg *et al.*, 1976; van Mansfeld *et al.*, 1978). Gene A plays a major role in the initiation and termination of ϕX174 RF→ss rolling-circle replication. Replication is initiated by Gene A, binding specifically to the (+) strand origin which is cleaved to provide a $3'$-^{OH} DNA primer for rolling circle replication and then forms a covalent linkage with the $5'$-^{P} group (Henry and Knipper, 1974; Eisenberg *et al.*, 1976; van Mansfeld *et al.*, 1978; Langeveld *et al.*, 1978; Ikeda *et al.*, 1979; Eisenberg and Kornberg, 1979). Rep helicase then interacts with Gene A to form a Rep-Gene A-DNA complex which remains associated as Rep helicase proceeds to unwind the supercoiled duplex DNA (Scott *et al.*, 1977; Kornberg *et al.*, 1978; Arai and Kornberg, 1981e). Replication proceeds in a rolling-circle manner from the $3'$-^{OH} primer, extended by DNA polymerase III holoenzyme. After replication recreates the (+) strand origin, Gene A performs a second cleavage which is coupled to ligation (circularisation) of the single-stranded viral DNA (Eisenberg *et al.*, 1977). During this process the viral stand is liberated where it either serves as a template for ss→RF replication or is packaged into phage particles (Eisenberg *et al.*, 1976; Eisenberg *et al.*, 1977).

The ϕX174 origin (30 bp, position 4299-4328), located within *gene A*, consists of several elements which include the binding site, recognition sequence and the cleavage site for Gene A protein, and an AT-rich spacer region (Figure 6.1) (Baas *et al.*, 1980). A consensus sequence of eight nucleotides for Gene A cleavage has been identified ($^{A}/_{T}ACT^{C}/_{T}GA^{T}/_{G}$) and cleavage occurs between the G and A nucleotides (van Mansfeld *et al.*, 1978, 1980, 1984; Heidekamp *et al.*, 1980). However, this consensus sequence alone is not sufficient for Gene A cleavage of ϕX174 RFI DNA (Heidekamp *et al.*, 1981). The binding site and the recognition sequence are also required (Fluit *et al.*, 1984a, 1984b). Gene A is able to cleave supercoiled RFI (Henry and Knippers, 1974; Ikeda *et al.*, 1976; Eisenberg *et al.*, 1977; Langeveld *et al.*, 1978) and single-stranded ϕX174 (Langeveld *et al.*, 1979; Eisenberg, 1980) but not relaxed duplex RF (Marians *et al.*, 1977).

(a)



(b)



(c)

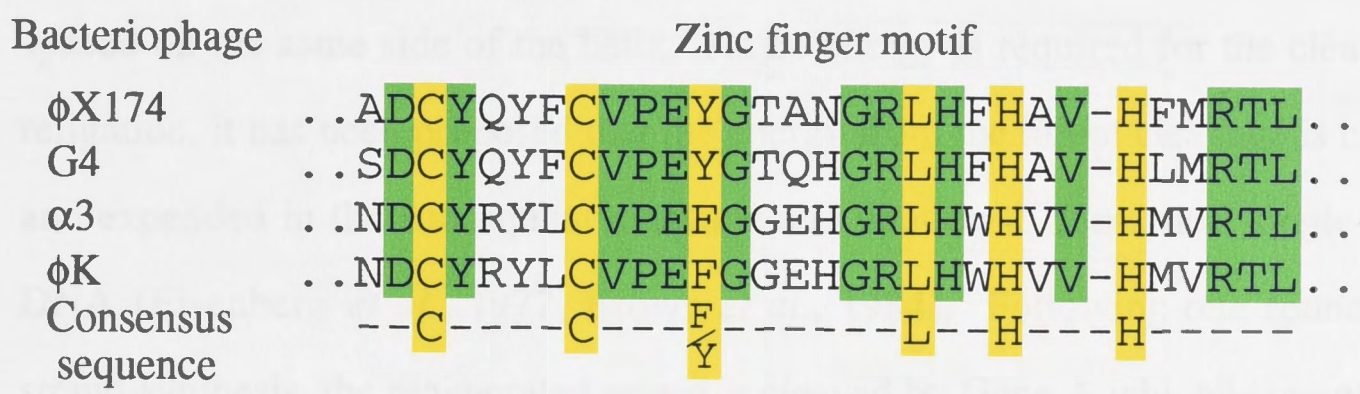


Figure 6.1

(a) A schematic of *gene A* showing *gene A** which is transcribed from an internal in-frame ATG codon and the positions of the origin and a possible zinc finger motif.

(b) The 30-bp origin of ϕX174 which contains the Gene A binding site, recognition sequence and the cleavage site as well as the AT-rich region.

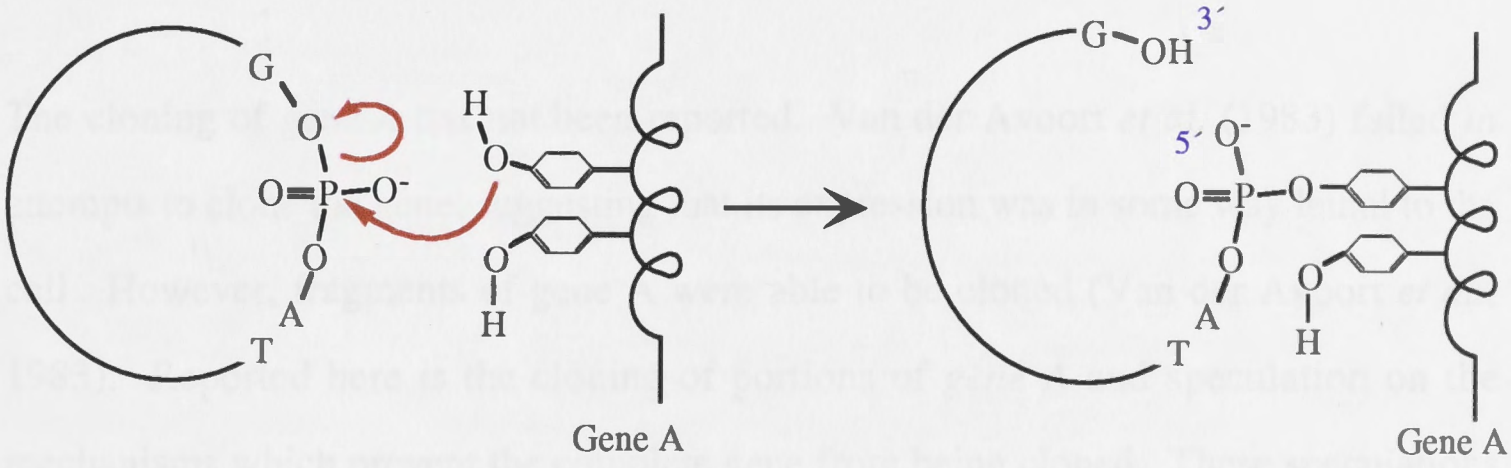
(c) A possible zinc finger motif identified in bacteriophages ϕX174, G4, α3 and ϕK compared with the concensus sequence of a cys_2-his_2 zinc finger. The yellow boxes indicate the homology with the concensus sequence and additional green boxes show homology between the sequences from the bacteriophages.

This has led to the belief that Gene A acts upon a single-stranded origin (Sanhueza and Eisenberg, 1985). It has been proposed that Gene A first binds to the origin inducing melting of the duplex DNA in the AT-rich spacer region and then the recognition region, providing a single-stranded sequence for cleavage by the protein (Hayashi *et al.*, 1988). Recently, a cys_2-his_2 zinc finger motif was identified in the amino acid sequences of ϕX174 Gene A and related proteins of bacteriophage G4, ϕK and α3 (Figure 6.2) (Kodaira *et al.*, 1989). However, the role of this putative zinc finger motif in Gene A function has not been elucidated.

Unlike Gene II protein of M13, Gene A forms a covalent linkage with the 5´-phosphate generated by the cleavage (Ikeda *et al.*, 1979; Eisenberg and Kornberg, 1979). The covalent linkage is formed via either of two tyrosine residues (343 and 347) of a repeating sequence (tyr-val-ala-lys-tyr-val-asn-lys) and the phosphate group of the adenosine at position 4306 (van Mansfeld *et al.*, 1986). It has been predicted that the two tyrosine residues reside in an α-helical portion of the protein and are closely spaced on the same side of the helix. As no energy is required for the cleavage and religation, it has been proposed that the energy from the initial cleavage is conserved and expended in the subsequent cleavage and ligation to form viral single-stranded DNA (Eisenberg *et al.*, 1977; Brown *et al.*, 1984). Following one round of viral strand synthesis, the regenerated origin is cleaved by Gene A which is coupled to the ligation of the 3´ and 5´ ends to form a single-stranded circular viral DNA. A mechanism for the cleavage and cleavage/ligation reactions has been proposed (Figure 6.2) (van Mansfeld *et al.*, 1986).

Another similarity between Gene A and Gene II is that a second protein is translated from an internal in-frame ATG start codon within the gene. In ϕX174 the protein is Gene A*, a 37-kDa protein that is identical to the C-terminus of Gene A. This protein is responsible for shutoff of host cell replication and the switch from RF→RF to RF→ss replication of ϕX174 by mechanisms which are not known (Linney and

Strand cleavage



Strand cleavage and ligation

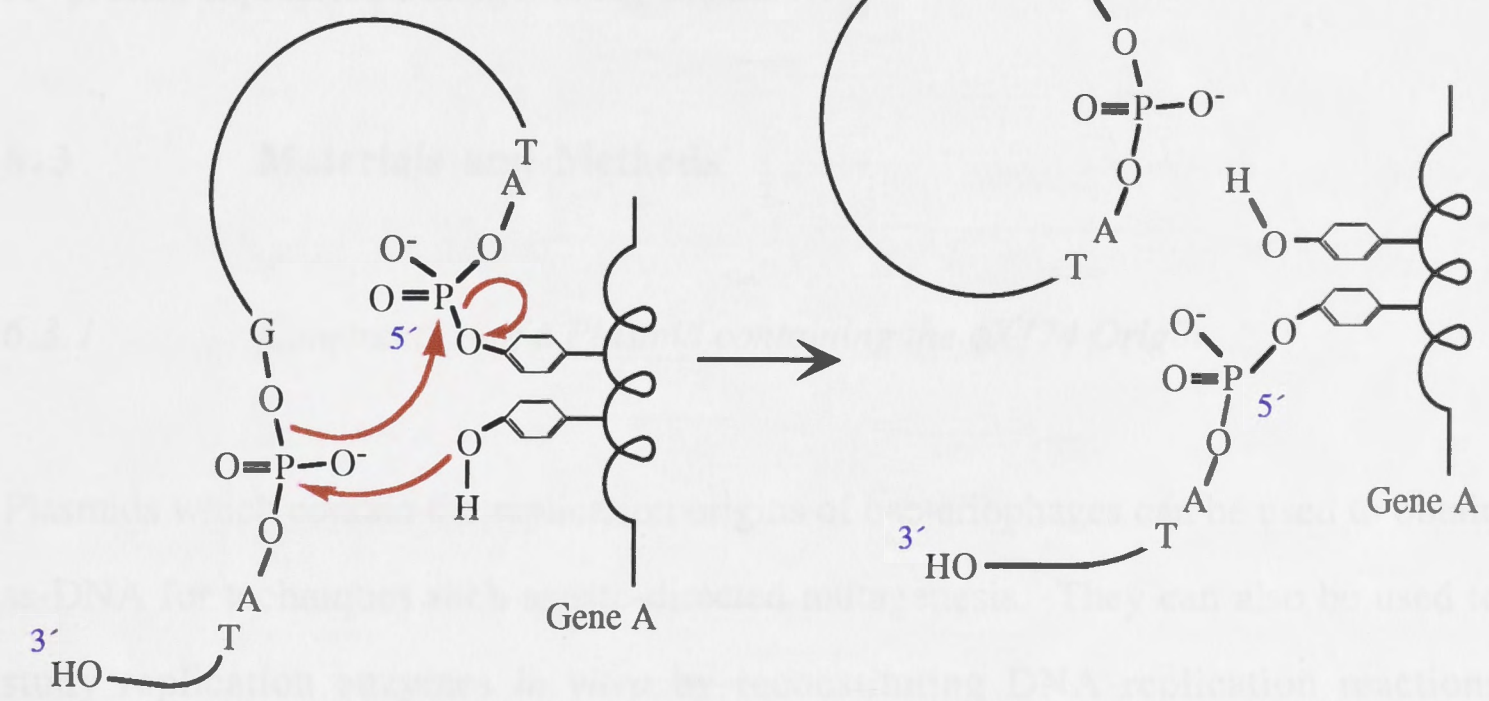


Figure 6.2

A model for the Gene A catalysed cleavage and cleavage/ligation reactions which initiate and terminate ϕX174 rolling-circle replication. The cleavage site is represented as -T-A-O-PO_2-O-G- and is part of the (+) strand origin.

Hayashi, 1973; Martin and Godson, 1975; Funk and Snover, 1976; Langeveld *et al.*, 1979; Langeveld *et al.*, 1981; Eisenberg and Ascarelli, 1981). Gene A* protein possesses the endonuclease and ligase properties of Gene A (Dubeau and Denhardt, 1981; Eisenberg and Finer, 1980; van der Ende *et al.*, 1981; van Mansfeld *et al.*, 1982), but cannot replace Gene A in phage RF→ss replication *in vitro*.

The cloning of *gene A* has not been reported. Van der Avoort *et al.* (1983) failed in attempts to clone the gene, suggesting that its expression was in some way lethal to the cell. However, fragments of gene A were able to be cloned (Van der Avoort *et al.*, 1983). Reported here is the cloning of portions of *gene A* and speculation on the mechanisms which prevent the complete gene from being cloned. These speculations are based on the identification of a putative *gene A** promoter and the effects of Gene A* protein expression during cloning of *gene A*.

6.3 Materials and Methods

6.3.1 Construction of a Plasmid containing the ϕX174 Origin

Plasmids which contain the replication origins of bacteriophages can be used to obtain ss-DNA for techniques such as site-directed mutagenesis. They can also be used to study replication enzymes *in vitro* by reconstituting DNA replication reactions (Eisenberg *et al.*, 1976). To study the functions of Gene A protein a plasmid containing the ϕX174 replication origin was constructed. This was achieved by inserting a dsDNA oligonucleotide adaptor, containing the ϕX174 origin, between *Bgl*I overhanging ends into pMA200 (Figure 2.1) linearised with *Bgl*I. The oligonucleotide was a mixture of 5′-TGGCAACTTGATATTAATAACACTATAGACCACGAA and 5′-GTGGTCTATAGTGTTATTAATATCAAGTTGCCATTC. Ampicillin resistant transformants of strain AN1459 were selected at 30°C and their plasmid DNAs were isolated on a small scale. Plasmids of the appropriate size (4185 bp) were digested

with *Bgl*I and *Taq*I endonucleases to confirm the presence of a 243-bp fragment containing the adaptor. Further confirmation of construction of this plasmid was achieved by nucleic acid sequencing. The plasmid was designated pCL772 (Figure 6.3).

6.3.2 Cloning of ϕX174 gene A Fragments

Several *gene A* fragments were generated by digesting ϕX174 DNA with *Mun*I-*Hae*III, *Mun*I-*Nru*I, *Mun*I-*Stu*I and *Mun*I-*Bss*HII (Figure 6.4). The *Bss*HII site of the *Mun*I-*Bss*HII fragment was made blunt using the Klenow enzyme. The fragments were isolated by agarose gel electrophoresis, then extracted from the gel using a QIAEX gel extraction kit. The *Mun*I-*Hae*III (268 bp), *Mun*I-*Nru*I (487 bp) and *Mun*I-*Stu*I (549 bp) fragments were ligated into pETMCSII (Figure 2.2) that had been digested with *Eco*RI and *Sma*I, such that they were in the reverse orientation with respect to the T7 promoter. The *Mun*I-*Bss*HII (1409 bp) fragment was inserted into pCL772 digested with *Eco*RI and *Sma*I. The constructs were transformed into strain AN1459 and ampicillin resistant transformants were selected. Plasmid DNAs were isolated on a small scale and the clones were confirmed by nucleic acid sequencing. The resulting plasmids were designated pCL886, 884, 883 and 889, respectively (Figure 6.4 and 6.5).

*6.3.3 Cloning of ϕX174 gene A**

*Gene A** was excised from ϕX174 using restriction endonucleases *Xho*I (made blunt using the Klenow enzyme) and *Stu*I. The plasmid pETMCSII was linearised with *Sma*I. The two fragments were then ligated and transformed into strain AN1459. Ampicillin transformants were selected and plasmid DNAs isolated. Transformants containing plasmids of the appropriate size (5725 bp) were confirmed by nucleic acid sequencing to contain gene A*. The resulting plasmid was designate pCL885

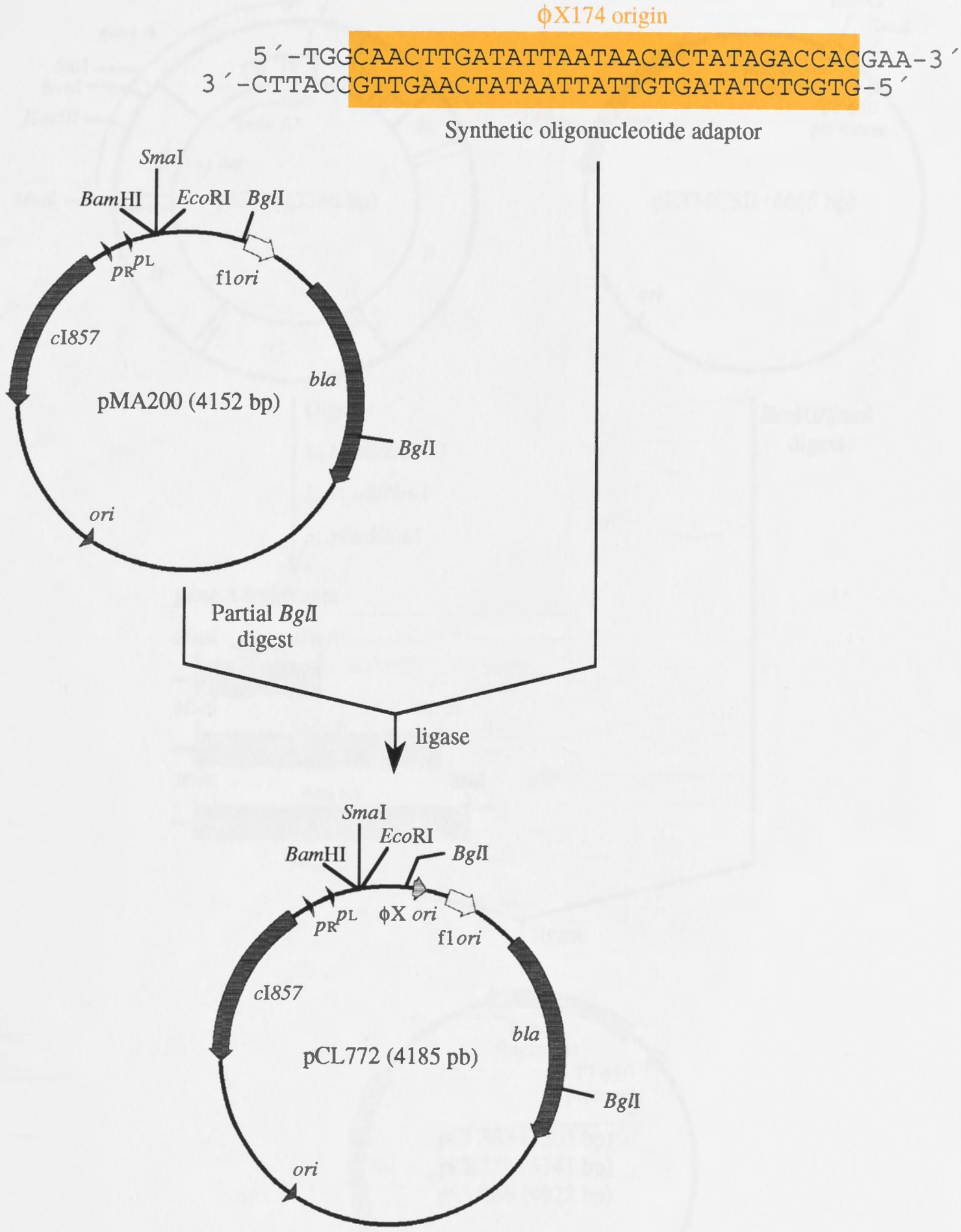


Figure 6.3

The scheme for the construction of pCL772, a vector that contains the ϕX174 origin. The vector was constructed by inserting a dsDNA synthetic oligonucleotide, containing the 30-bp origin of ϕX174, into one of the *Bgl*I sites of pMA200.

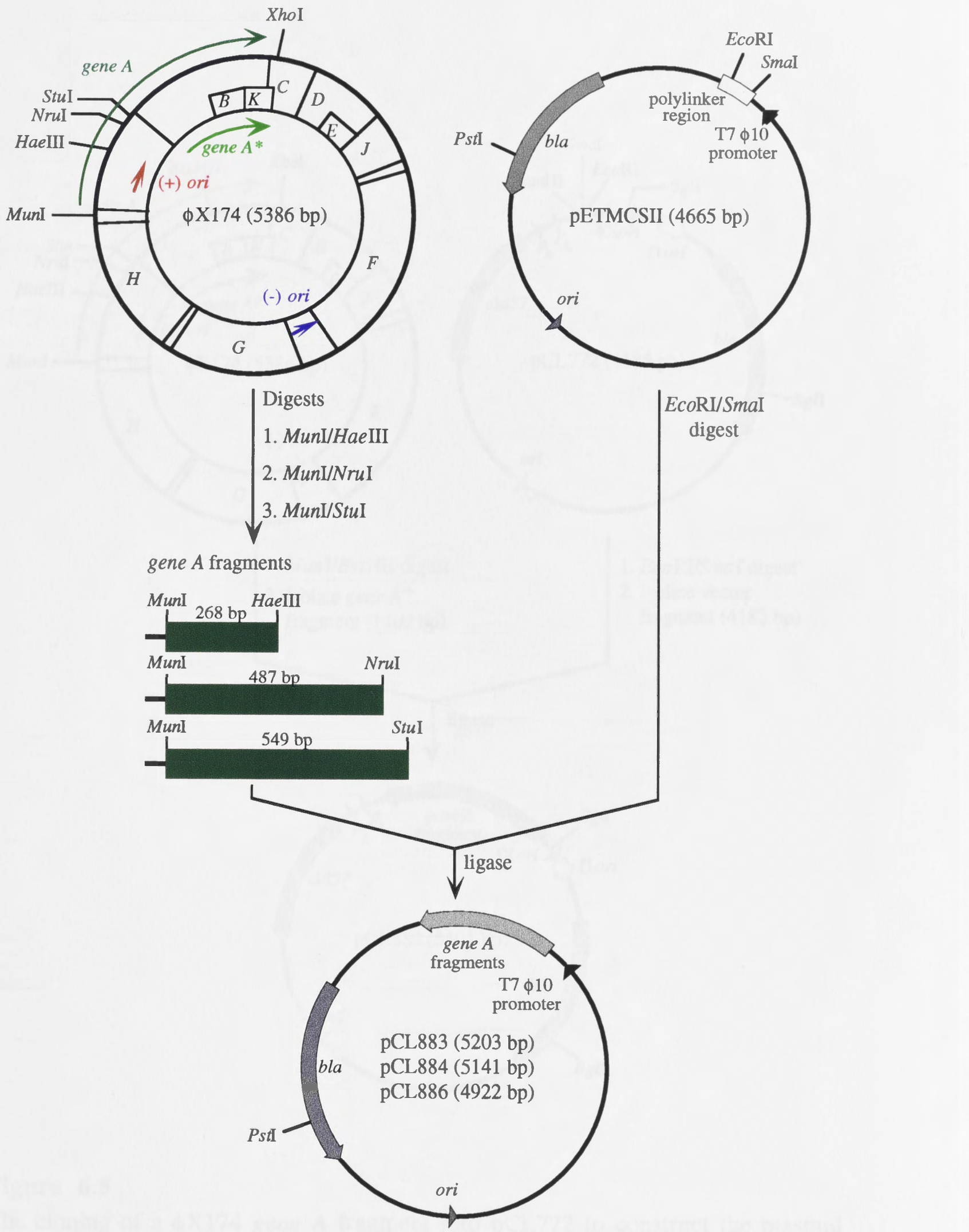


Figure 6.4

The scheme for the cloning of ϕX174 *gene A* fragments into pETMCSII to create the plasmids pCL883, 884 and 886.

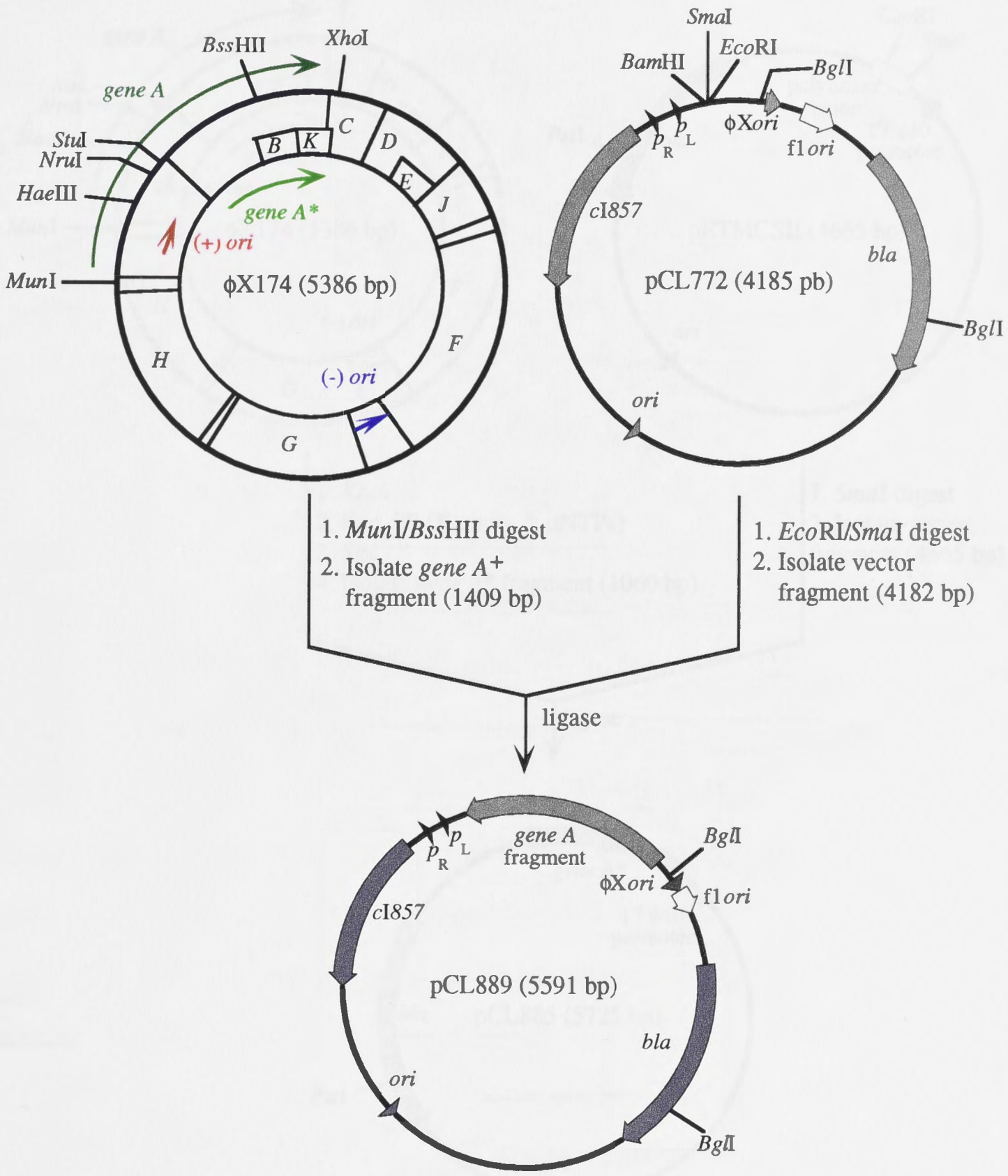
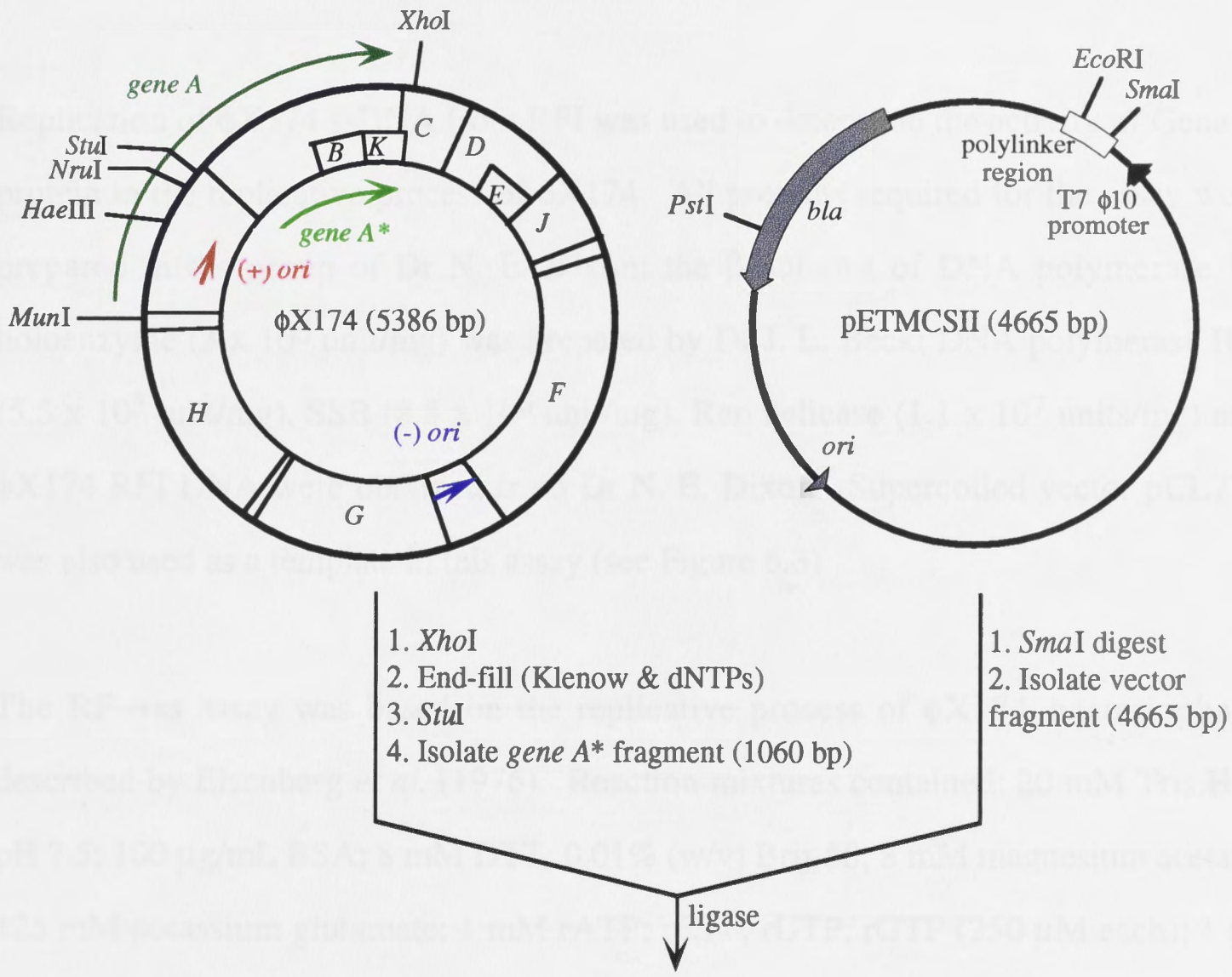


Figure 6.5

The cloning of a ϕX174 *gene A* fragment into pCL772 to construct the plasmid pCL889. ϕX174 was digested with *Mun*I and *Bss*HII and the *gene A* fragment inserted between the *Sma*I and *Eco*RI sites of pCL772 to create pCL889.



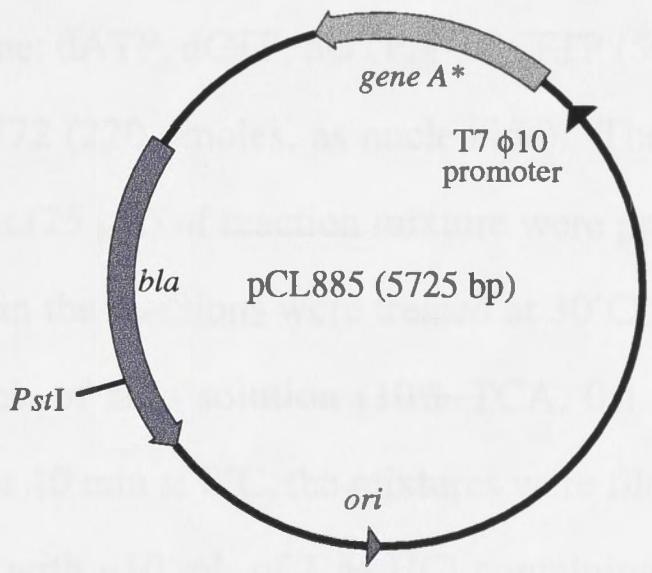


Figure 6.6

The scheme for cloning of ϕX174 *gene A** into pETMCSII. ϕX174 was digested with *Xho*I and the ends were made blunt, then digested with *Stu*I. A fragment containing *gene A** was isolated and inserted into the *Sma*I site of pETMCSII to create pCL885.

(Figure 6.6).

6.3.4 RF→ss Replication Assay for Gene A Protein

Replication of ϕX174 ssDNA from RFI was used to determine the activity of Gene A protein in the replicative process of ϕX174. All proteins required for the assay were prepared in the group of Dr N. E. Dixon: the β subunit of DNA polymerase III holoenzyme (3 x 10^6 unit/mg) was prepared by Dr J. L. Beck; DNA polymerase III* (5.5 x 10^5 unit/mg), SSB (8.8 x 10^6 unit/mg), Rep helicase (1.1 x 10^7 units/mg) and ϕX174 RFI DNA were obtained from Dr N. E. Dixon. Supercoiled vector pCL772 was also used as a template in this assay (see Figure 6.3)

The RF→ss assay was based on the replicative process of ϕX174 bacteriophage described by Eisenberg *et al.* (1976). Reaction mixtures contained: 20 mM Tris.HCl pH 7.5; 100 μg/mL BSA; 8 mM DTT; 0.01% (w/v) Brij-58; 8 mM magnesium acetate; 125 mM potassium glutamate; 1 mM rATP; rCTP, rUTP, rGTP (250 μM each); 1 μg of SSB; Rep helicase; 100 ng of DNA polymerase III*; 26 ng of β subunit of DNA polymerase III holoenzyme; dATP, dCTP, dGTP, [^{3}H]dTTP (50 μM each) and ϕX174 RFI or supercoiled pCL772 (220 pmoles, as nucleotide). The reaction mixture was prepared at 0°C. Aliquots (25 μL) of reaction mixture were prepared at 0°C and after addition of Gene A protein the reactions were treated at 30°C for 20 min. Cooling to 0°C and addition of 1 mL of stop solution (10% TCA, 0.1 M $NaPP_i$) terminated replication of DNA. After 10 min at 0°C, the mixtures were filtered through Whatman GF/C filters and washed with ~10 mL of 1 M HCl containing 0.1 M $NaPP_i$. Filters were then washed with ethanol, dried and counted in a Beckman LS 6000 IC liquid scintillation counter. Ready-Safe liquid scintillation cocktail was purchased from Beckman. Activity of Gene A protein in DNA replication was calculated in units (U), where one unit of replication activity denotes incorporation of one pmol of nucleotide into product per minute at 30°C.

6.4 Results and Discussion

6.4.1 RF→ss Replication of pCL772

The vector pCL772 was constructed as an alternative to ϕX174 RFI DNA for replication reconstitution assays to further study the functions of Gene A. pCL772 was used in the RF→ss replication assay (Section 6.3.4) using purified enzymes and the level of replication compared favourably to that obtained using ϕX174 RFI in the same assay. When Gene A (0.052 mg/mL, obtained from Dr N. E. Dixon) was titrated into the assay similar levels of replication were obtained from pCL772 and ϕX174 templates (Figure 6.7). Similar results were also obtained by Brown *et al.* (1983) who used the 30-bp ϕX174 (+) strand origin inserted into pBR322 as template. They also reported that the orientation of the origin had an effect on both the rate and yield of single-stranded product.

6.4.2 Cloning of ϕX174 gene A

The cloning of ϕX174 *gene A* has not been reported. Van der Avoort *et al.* (1983) were unable to clone gene A and reported that the gene may be lethal to the cell. Several attempts to clone the intact *gene A* as a *Mun*I-*Xho*I fragment (Figure 6.8) into vectors (pBluescript SK and KS, pETMCSII, pMTL23P, pPL450 and pPL451) in either orientation were unsuccessful. However, it was possible to clone several *gene A* fragments and *gene A**. The cloned fragments covered the entire gene (Figure 6.8) which suggests that it is the function of the expressed protein rather than the DNA sequence itself which is responsible for its lethal effect on cells. Attempts were also made to insert a his_6-tail at the start of *gene A* in ϕX174, to assist in purification of his_6-Gene A protein from infected cells. However, several attempts to do this proved unsuccessful.

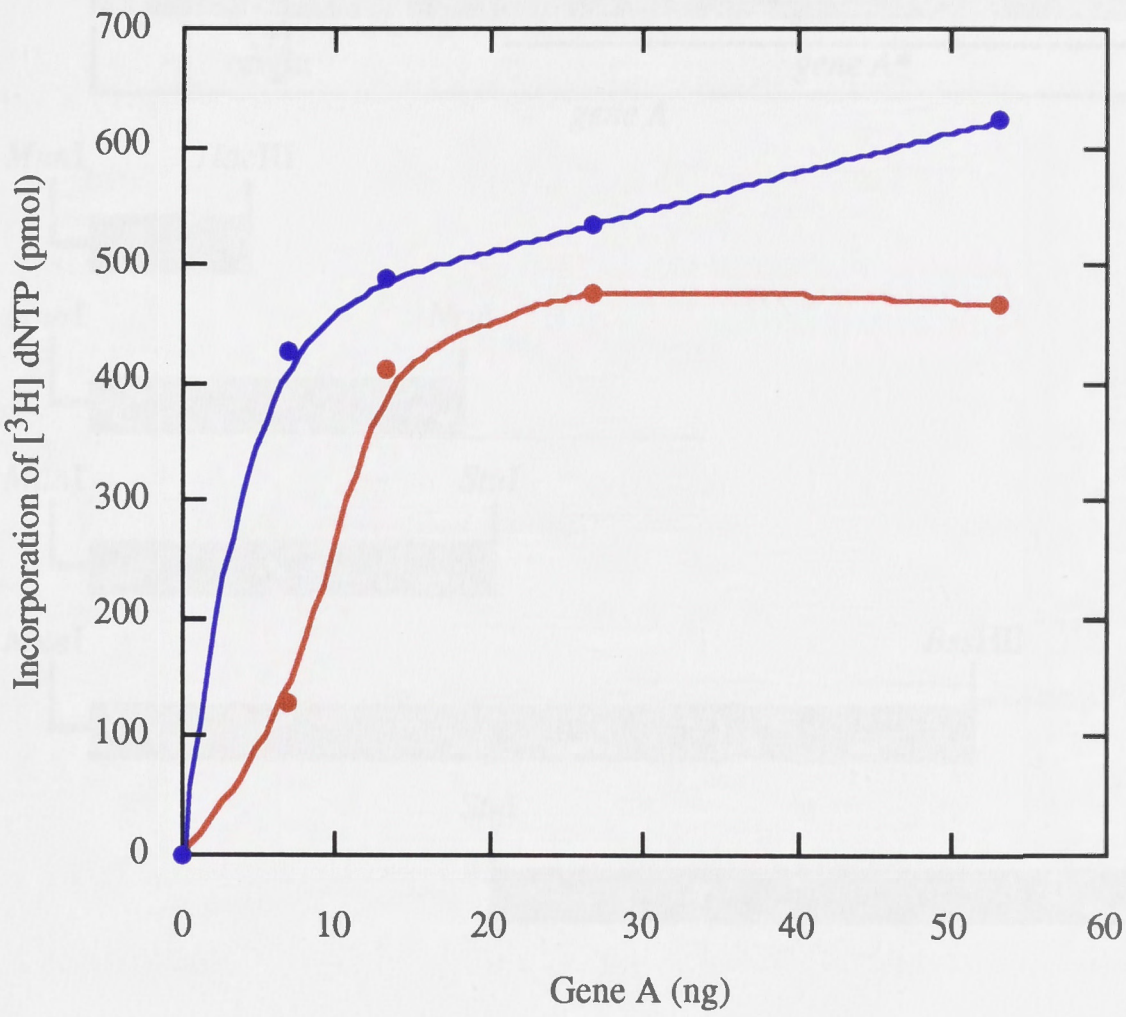


Figure 6.7

RF→ss replication of pCL772 (blue) and ϕX174 (red) using varying amounts of Gene A. The reactions were carried out as described in Section 6.3.3.

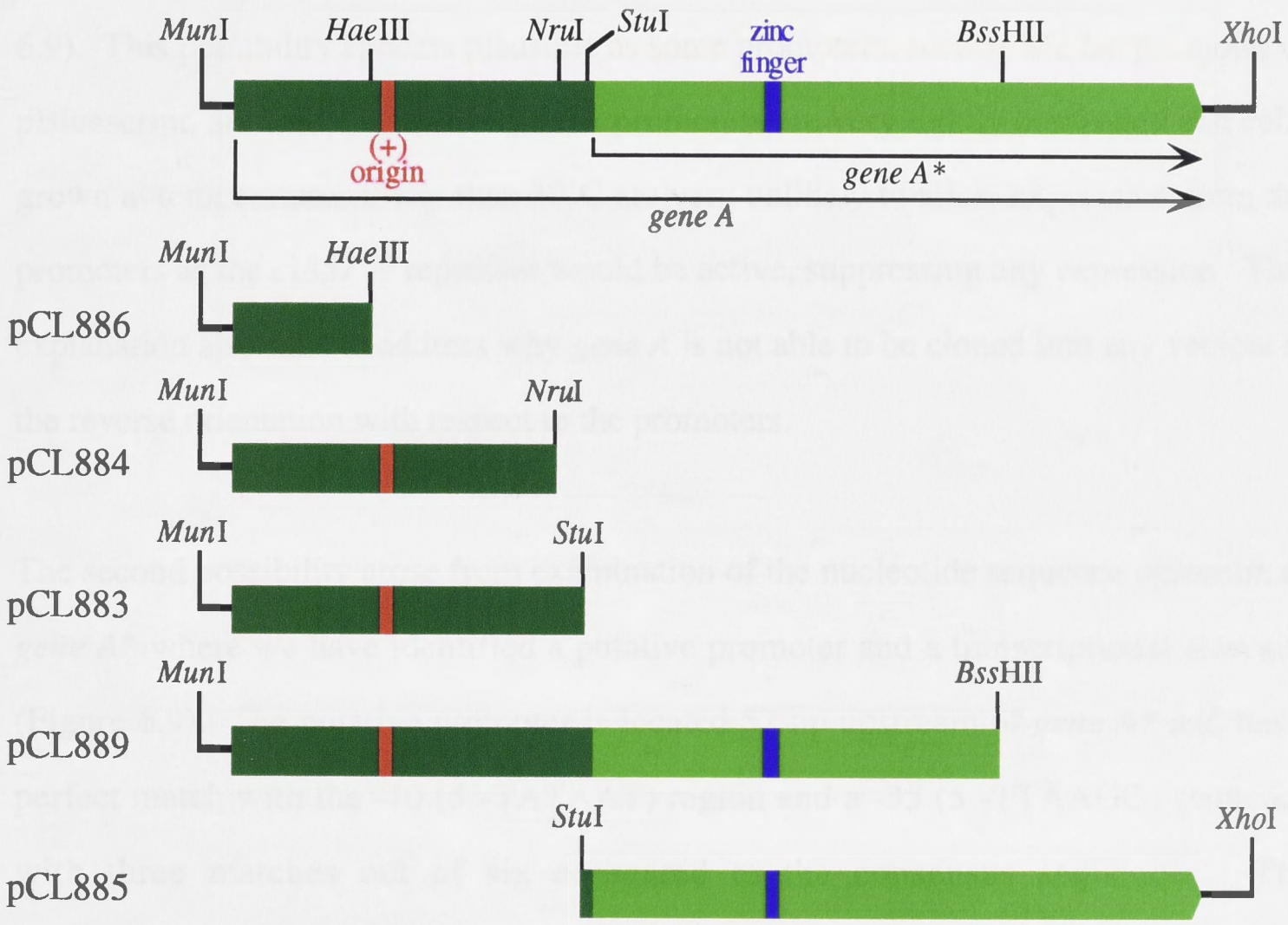


Figure 6.8

Gene A fragments that have been successfully cloned and the plasmids which contain them.

There are two explanations which may provide an insight as to why *gene A* is unable to be cloned. Firstly, it is possible that the vector promoters are leaky and lead to low-level expression of *gene A* and *gene A** during cloning. The Gene A and Gene A* produced are able then to cleave the ϕX174 origin which is located within the first third of *gene A*. Plasmids that are cleaved are then unable to be replicated *in vivo*, resulting in a lack of antibiotic resistance and, thus, no viable *gene A* clones are obtained (Figure 6.9). This possibility appears plausible as some promoters, such as the *lac* promoter in pBluescript, are leaky. However, the λ promoters are very tightly controlled and cells grown at temperatures lower than 30°C are very unlikely to allow expression from the promoters as the *cI857*$^{-ts}$ repressor would be active, suppressing any expression. This explanation also fails to address why *gene A* is not able to be cloned into any vectors in the reverse orientation with respect to the promoters.

The second possibility arose from examination of the nucleotide sequence upstream of *gene A** where we have identified a putative promoter and a transcriptional start site (Figure 6.9). The putative promoter is located 57 bp upstream of *gene A** and has a perfect match with the -10 (5´-TATAAT) region and a -35 (5´-TTAAGG) sequence with three matches out of six compared to the consensus sequences. The transcriptional start (5´-CAA) has two of three matches and is optimally spaced from the -10 region. Thus, it is possible that following transformation of gene A clones, *gene A** is being expressed constitutively from its own promoter. Gene A* is then able to cleave the ϕX174 origin located in *gene A*, resulting in the inability to obtain viable *gene A* clones as described above (Figure 6.9). If this postulation is correct, then it would be impossible to clone native *gene A* in either orientation. However, there has been no report of *gene A** mRNA transcripts in studies of *in vivo* and *in vitro* ϕX174 transcription. Three major transcripts have been identified and are located at the beginning of genes A, B and D (Smith and Sinsheimer, 1976a, 1976b, 1976c; Axelrod, 1976). However, several minor transcripts were observed in these studies (Smith and Sinsheimer, 1976a) and not all have been characterised. One of

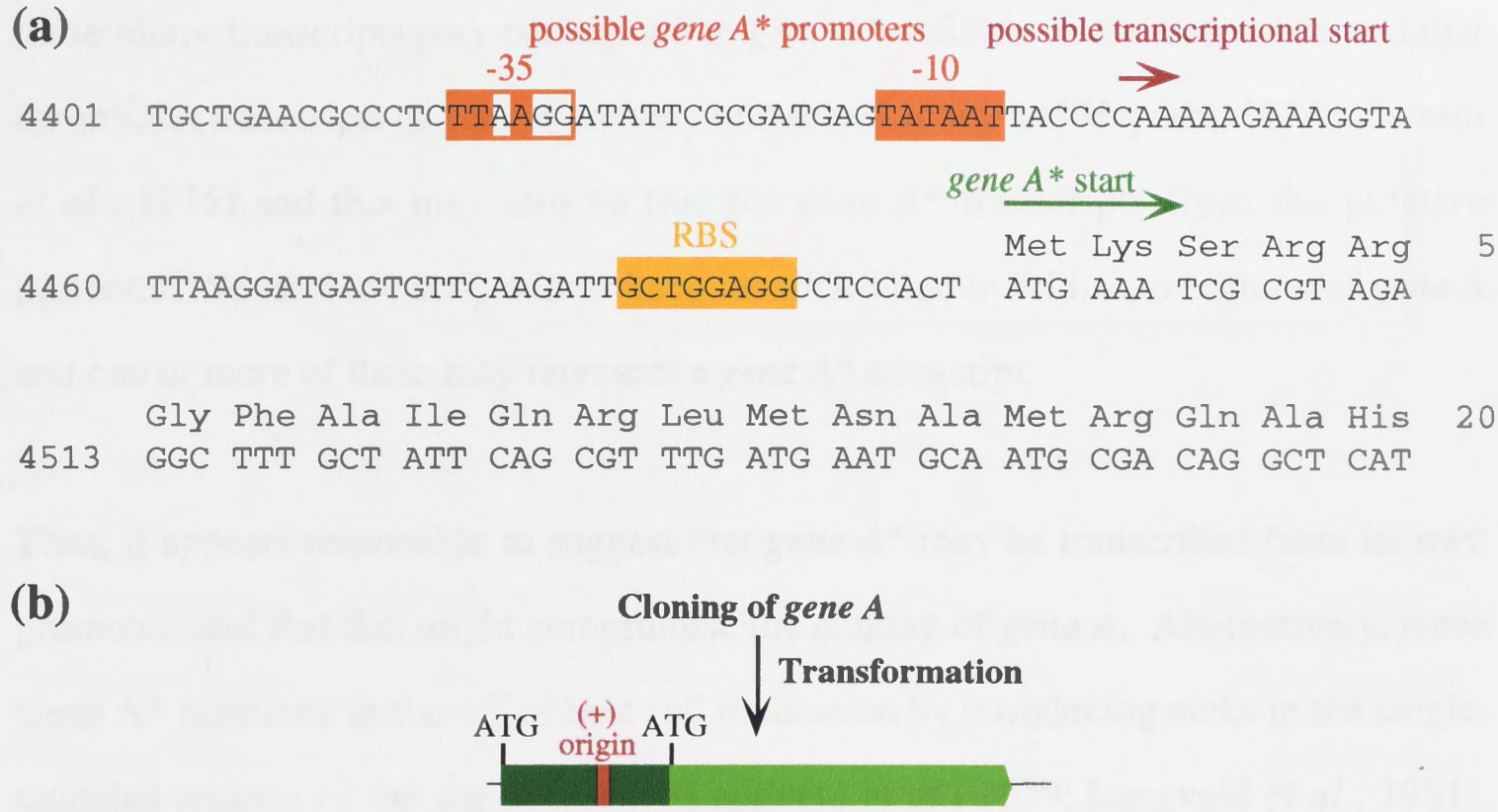


Figure 6.9

(a) The location of a putative promoter upstream of *gene A** of ϕX174 showing the positions of the -10 and -35 sites, and the transcriptional start.

(b) A scheme showing possible mechanisms which prevent the cloning of intact *gene A*.

these minor transcripts may correspond to *gene A** mRNA. It has been observed that the mRNA transcripts of *gene A* are very unstable (Linney and Hayashi, 1974; Hayashi *et al.*, 1976) and this may also be true for *gene A** transcripts from the putative promoter. Several transcripts have been identified that hybridise to regions of *gene A* and one or more of these may represent a *gene A** transcript.

Thus, it appears reasonable to suggest that *gene A** may be transcribed from its own promoter, and that this might compromise the cloning of *gene A*. Alternatively, since Gene A* functions in shutoff of host cell replication by introducing nicks in the single-stranded regions of the chromosome (Langveld *et al.*, 1979; Langveld *et al.*, 1981), this may have a lethal effect on cell growth during transformation of *gene A*$^+$ plasmids (Figure 6.9).

Site-directed mutagenesis has been employed to overcome the inability of cloning *gene A* and two experiments are underway. The first is to remove the *gene A** start (M182L) to eliminate constitutive expression of this gene during cloning. In the second experiment, the cleavage site of the ϕX174 (+) strand origin will be changed (D118N) to ensure that low level *gene A* expression during cloning does not result in cleavage of the origin (located within the gene). The double mutant will also be constructed.

6.5 Conclusion

In this chapter the construction of a plasmid containing the 30-bp ϕX174 (+) strand origin (pCL772) was described. Using purified enzymes, RF→ss replication of pCL772 was observed and the level of replication was comparable to that obtained from ϕX174 RFI DNA. This plasmid serves two purposes; to study ϕX174 RF→ss replication and to obtain ssDNA *in vitro* by RF→ss replication using purified enzymes.

Several fragments of *gene A* were cloned as well as *gene A**. However, attempts to clone *gene A* were unsuccessful. Investigation of the nucleic acid sequence of *gene A* revealed a putative promoter upstream of the start of *gene A**. The promoter was located 57 bp upstream of the *gene A** start, containing -10, -35 and mRNA start sites which show high homology to the consensus sequences of these components, and were spaced in accordance with a functional promoter. During the cloning of *gene A*, *gene A** may be expressed constitutively from its own promoter, compromising the experiments by either having a lethal effect on cell growth by shutoff of cell replication or cleavage of the ϕX174 (+) strand origin, affecting plasmid replication (Figure 6.9). As a result cloning of intact *gene A* was not possible. Although there has been no report of mRNA transcripts of *gene A**, several minor transcripts have been identified but not characterised (Smith and Sinsheimer, 1976a). Also, a number of transcripts have been identified which hybridised to *gene A* (Hayashi *et al.*, 1976). These transcripts were very unstable and degraded quickly (Linney and Hayashi, 1974; Hayashi *et al.*, 1976). Site-directed mutagenesis experiments are being carried out to remove the cleavage site of the ϕX174 (+) strand origin located in *gene A* and to remove the start codon of *gene A** in an attempt to clone *gene A*.

CHAPTER 7

CONCLUSION

The process of initiation of DNA replication of the *E. coli* chromosome involves several proteins (DnaA, DnaB, DnaC as well as other proteins) which work in concerted fashion to open the duplex DNA and load the replication machinery. DnaA protein is the key protein involved in initiation of DNA replication, and also play important roles in the regulation, timing and synchrony of this process. *E. coli* single-stranded DNA bacteriophages, M13 and ϕX174, rely on their host for the supply of all the replication machinery with the exception of the replication initiator which is an endonuclease encoded by the phage. These replication initiator [Gene II (M13) and Gene A (ϕX174)] proteins initiate and terminate replication by cleavage and ligation of replication intermediates. The process of initiation of bacteriophage DNA replication is very different from the way it is preformed in replication of the *E. coli* chromosome by DnaA.

Studies on three replication initiator (DnaA, Gene II and Gene A) proteins was the focus of this research. All these proteins have presented problems associated with obtaining them in large quantities for structural and functional biochemical studies. They are insoluble and unstable, and thus difficult to purify. The reported methods for their purification are tedious and as a result, very low yields were obtained. High copy number expression vectors have been used to increase the cellular levels of these proteins, and new purification procedures have been developed to obtain them in large quantities.

The overproduction of DnaA was improved by Dr N.E. Dixon and other members of his research group by replacing the GUG start codon with AUG, replacing the natural RBS with a synthetic RBS perfectly complementary to the 3´ end of the *E. coli* 16S rRNA and inserting the gene into an expression vector under the control of strong λ promoters. This resulted in very high level of expression which represented approximately 84 000 molecules per cell. The spacing between the RBS and the ATG start codon was investigated to determine the effect this spacing had on gene

expression. Optimal expression of DnaA and Tus proteins was observed when these elements was separated by 7 nucleotides. However, the level of expression varied only 2-fold with spacings from 5 to 11 nucleotides. Gene expression was reduced 8-10 fold when the RBS and the ATG start codon were separated by 16 nucleotides.

Purification of DnaA has been hampered due to the monomer aggregation and insolubility at low ionic strength. The presence of ATP and $MgCl_2$ were found to stabilise the protein at moderate ionic strengths. This has enabled a method for large-scale purification to be developed which does not require the use of protein denaturants. The procedure involves cell lysis by a combination of lysozyme treatment and freeze/thawing, ammonium sulfate fractionation and chromatography on cellulose phosphate and Sephadex G-200, resulting in highly pure, fully active, monomeric DnaA in an overall yield of 25%. Replication of M13 A-site ssDNA which contains a single DnaA binding site as part of a hairpin structure, required less than two molecules of DnaA per template molecule.

Sedimentation equilibrium studies confirmed that the purified DnaA was monomeric in solution. The monomeric molecular weight of DnaA was accurately determined using electrospray mass spectrometry to be 52 422.5 Da which was within 2.8 Da of the theroetical mass (52 419.7 Da). A second component 79.9 Da larger was also observed which may suggest that DnaA is either phosphorylated or sulfated, although the former is more likely as several eukaryotic replication proteins are phosphorylated and dephosphorylated to regulate specific events in the cell cycle. Experiments to phosphorylate DnaA *in vitro* with ^{32}P using a crude cell extract from a *dnaA* mutant strain produced labelled protein an autoradiograph in a position that corresponded to that of DnaA, providing further evidence of phosphorylation. Digestion of native DnaA with trypsin and analysis of the peptide fragments by electrospray and tandem mass spectrometry have shown that the site of putative phosphorylation resides in the first 107 residues. This region was found to contain a disulfide linkage between

residues 8 and 66 which prevented the native protein from being digested by trypsin. The site of putative phosphorylation has not yet been determined precisely, although this is still under active investigation.

UV spectral simulations were used to determine the number of ATP molecules bound to DnaA in solution. Simulated UV spectra containing one, two or three ATP molecules were compared with the observed DnaA spectrum. It was not possible to accurately determine the levels of ATP bound to DnaA as there appears to be absorbance at 260 nm from a chromophore other than ATP. It was possible to simulate accurately the UV spectrum of *E. coli* prolidase, used as a model for a protein with no chromophoric prosthetic group.

M13 Gene II and Gene X proteins were overproduced to high levels following insertion of the *gene II* into expression vectors under the control of strong λ promoters. The overproduction was further improved (10-fold) in T7 promoter vectors. However, the overproduced protein was insoluble. The protein required either 8 M urea or 6 M guanidine.HCl to be completely soluble in solution. Following screening of buffers, pH and surfactants, it was found that Gene II was soluble, when diluted from a denatured state, in 25 mM Tris.HCl pH 9.0-9.5 containing 1 mM EDTA, 14 mM β-mercaptoethanol and 1 M urea.

A procedure for purifying Gene II protein was developed. Cells were lysed in a French press and the insoluble portion washed several times with 4 M urea, which removed all the contaminating Gene X protein from the solution. Gene II, insoluble in 4 M urea, was treated with 8 M urea and then diluted in 25 mM Tris.HCl pH 9.0-9.5 containing 1 mM EDTA, 14 mM β-mercaptoethanol. Approximately 30-60 mg of protein, judged to be >95% pure, was obtained from 1 L of cell culture which represented 20-30% of the total cell protein. The protein was found to be active in nicking M13 RFI DNA at levels above 5 molecules of protein per template circle.

However, the protein was only active when diluted into the assay from a solution containing 8 M, urea which suggests that Gene II in an solution in 25 mM Tris.HCl pH 9.0-9.5, 1 mM EDTA, 14 mM β-mercaptoethanol, 1 M urea must be incorrectly folded. Gene II was not active in RF replication of M13, although we have speculated that an unknown factor removed during the purification is also required for this process. It has been reported by others that replication activity diminished in highly purified preparations.

Proteins tagged with a poly-histidine tail can often be purified in a one-step procedure using metal chelate [Ni(II)-NTA] chromatography. Expression vectors were constructed to tag protein at the N-terminus with six histidine residues and the level of overproduction of his_6-Gene II was similar to that obtained with the wild type protein. His_6-Gene II was bound to the Ni(II)-NTA resin under denaturing conditions, then the urea was diluted gradually from 8 to 1 M over several hours. Pure, soluble protein was eluted from the column. However, the protein was inactive in relaxation of M13 RFI DNA.

There has been no reported cloning of ϕX174 *gene A*. Several fragments of *gene A* and *gene A** were cloned but attempts to clone the intact gene were unsuccessful. Investigation of the nucleic acid sequence of *gene A* revealed a putative promoter upstream of the start of *gene A**. The promoter was located 57 bp upstream of the *gene A** start, and contained -10 and -35 regions and an mRNA start site which showed homology to the consensus sequences of these components, and were spaced in accordance with a functional promoter. During the cloning of *gene A*, *gene A** may be expressed constitutively from its own promoter, thereby compromising the experiments by either having a lethal effect on cell growth by shutoff of chromosomal replication or cleavage of the ϕX174 (+) strand origin, affecting plasmid replication. As a result, cloning of intact *gene A* was not possible.

APPENDIX 1

Fast Screen

A list of buffers routinely used for a "fast screen" in initial crystallisation experiments. The *pH of cold solutions was estimated using pH paper. Abbreviations used in this table are: AmAc - ammonium acetate, AmS - ammonium sulfate, K Pi (dibasic) - K_2HPO_4/KH_2PO_4, NaAc - sodium acetate, Na formate - sodium formate, Na tartrate - sodium tartrate.

1	0.2 M $CaCl_2$, 0.1 M NaAC pH 4.7, 30% MPD	25	0.1 M HEPES pH 7.5, 1 M NaAc
2	0.4 M Na tartrate pH 6-6.5	26	0.2 M AmAc, 0.1 M citrate pH 5.5, 30% MPD
3	0.4 M K P_i (dibasic) pH 8.5	27	0.2 M NaCit, 0.1 M HEPES pH 7.5, 20% isopropanol
4	0.1 M Tris pH 8.5	28	0.2 M NaAc, 0.1 M cacodylate pH 6.5, 30% PEG 8K
5	0.2 M NaCit, 0.1 M HEPES pH 7.5, 40% MPD	29	0.1 M HEPES pH 7.5, 0.8 M Na tartrate
6	0.2 M $MgCl_2$, 0.1 M Tris pH 8.5,15% PEG 3.4K	30	0.2 M AmS, 30% PEG 8K
7	0.1 M cacodylate pH 6.5, 1.3 M NaAc	31	0.2 M AmS, 30% PEG 3.4 K, *pH 5
8	0.2 M NaCit, 0.1 M cacodylate pH 6.5, 30% isopropanol	32	2 M AmS *pH5
9	0.2 M AmAc, 0.1 M citrate pH 5.5, 15% PEG 3.4K	33	4 M Na formate *pH 6-6.5
10	0.2 M AmAc, 0.1 M NaAc pH 4.7, 30% PEG 3.4K	34	0.1 M NaAc pH 4.7, 2 M Na formate
11	0.1 M Citrate pH 5.5, 1 M K P_i (dibasic)	35	0.1 M HEPES pH 7.5, 1.8 M Na tartrate
12	0.2 M $MgCl_2$, 0.1 M HEPES pH 7.5, 30% isopropanol	36	0.1 M Tris pH 8.5, 8% PEG 8K
13	0.2 M NaCit, 0.1 M Tris pH 8.5, 30% PEG 400	37	0.1 M NaAc pH 4.7, 8% PEG 3.4K
14	0.2 M $CaCl_2$, 0.1 M HEPES pH 7.5, 28% PEG 400	38	0.1 M HEPES pH 7.5, 1.35 M NaCit
15	0.2 M AmS, 0.1 M cacodylate pH 6.5, 30% PEG 8K	39	0.1 M HEPES pH 7.5, 2% PEG 400, 2 M AmS
16	0.1 M HEPES pH 7.5, 1.5 M Li_2SO_4	40	0.1 M Citrate pH 5.5, 20% isopropanol, 20% PEG 3.4K
17	0.2 M Li_2SO_4,0.1 M Tris pH 8.5	41	0.1 M HEPES pH 7.5, 10% isopropanol, 20% PEG 3.4K
18	0.2 M MgAc, 0.1 M cacodylate pH 6.5, 20% PEG 8K	42	0.05 M K P_i (dibasic), 20% PEG 8K, *pH 8.5-9
19	0.2 M AmAc, 0.1 M Tris pH 8.5, 30% isopropanol	43	30% PEG 2000 *pH 5.5
20	0.2 M AmAc, 0.1 M NaAc pH 4.7, 30% PEG 3.4K	44	0.2 M Mg formate
21	0.2 M MgAc, 0.1 M cacodylate pH 6.5, 30% MPD	45	0.2 M ZnAc, 0.1 M cacodylate pH 6.5, 18% PEG 8K
22	0.2 M NaAc, 0.1 M Tris pH 8.5, 30% PEG 3.4 K	46	0.2 M CaAc,0.1 M cacodylate pH 6.5, 18% PEG 8K
23	0.2 M $MgCl_2$, 0.1 M HEPES pH 7.5, 30% PEG 400	47	2 M Na/K P_i pH 8
24	0.2 M $CaCl_2$, 0.1 M NaAc pH 4.7, 20% isopropanol	48	2 M Na/K P_i pH 6

Medium Screen

Buffers used in the "medium screen" for crystallisation experiments. Abbreviations used in this table are: AmAc - ammonium acetate, AmS - ammonium sulfate, K Pi (dibasic) - K_2HPO_4/KH_2PO_4, NaAc - sodium acetate, Na formate - sodium formate, Na tartrate - sodium tartrate.

1	0.5 M NaCit pH 6	29	2.3 M Phosphate pH 7
2	0.1 M K P_i (dibasic) pH 8, 2 M AmS	30	2.6 M Phosphate pH 7
3	0.1 M Tris pH 8, 2 M AmS	31	0.03 M MES pH 6, 0.01 M $MgSO_4$, 20% PEG 8K
4	2 M Phosphate	32	0.03 M MES pH 6, 0.01 M $MgSO_4$, 1 M AmS
5	3 M Phosphate	33	0.1 M Phosphate pH 7, 1.2 M AmS
6	4 M Phosphate	34	0.03 M MES pH 6, 1 M Na_2SO_4, 20% PEG 8K
7	1 M NaCit pH 6	35	0.03 M MES pH 6, 0.4 M Na_2SO_4, 20% PEG 8K
8	1.5 M NaCit pH 6	36	0.03 M MES pH 6, 0.1 M Na_2SO_4, 20% PEG 8K
9	0.02 M Tris pH 8, 20% PEG 3.4K	37	0.03 M MES pH 6, 0.5 M $ZnSO_4$, 20% PEG 8K
10	0.02 M Tris pH 8, 20% PEG 8K	38	0.03 M MES pH 6, 0.01 M $ZnSO_4$, 1 M AmS
11	0.02 M Tris pH 8, 20% PEG 20K	39	0.03 M MES pH 6, 0.4 M $CuSO_4$, 20% PEG 8K
12	0.75 M NaCit pH 6	40	0.03 M MES pH 6, 0.01 M $CuSO_4$, 20% PEG 8K
13	0.5 M NaCit pH 6, 20% MPD	41	0.03 M MES pH 6, 1 M Li_2SO_4, 20% PEG 8K
14	0.5 M NaCit pH 6, 60% MPD	42	0.1 M Citrate pH 6, 1 M Li_2SO_4, 20% MPD
15	0.1 M Phosphate pH 7, 1 M AmS	43	0.2 M MES pH 6, 0.01 M $MgCl_2$, 20% MDP
16	0.1 M Phosphate pH 7, 1.5 M AmS	44	0.15 M MES pH 6, 0.01 M $MgCl_2$, 60% MPD
17	0.03 M MES pH 6, 1 M NaCl, 20% PEG 3.4K	45	0.03 M MES pH 6, 20% PEG 3.4K
18	0.03 M MES pH 6, 1 M NaCl, 20% PEG 8K	46	0.03 M MES pH 6, 20% PEG 8K
19	0.03 M MES pH 6, 1 M NaCl, 20% PEG 20K	47	0.03 M MES pH 6, 20% PEG 20K
20	0.03 M MES pH 6, 0.5 M NaCl, 20% PEG 3.4K	48	0.03 M MES pH 6, 30% PEG 3.4K
21	0.03 M MES pH 6, 0.5 M NaCl, 20% PEG 8K	49	0.03 M MES pH 6, 30% PEG 8K
22	0.03 M MES pH 6, 0.5 M NaCl, 20% PEG 20K	50	0.03 M MES pH 6, 30% PEG 20K
23	0.02 M Tris pH 8, 30% PEG 3.4K	51	0.1 M NaCit pH 6, 2 M AmS
24	0.02 M Tris pH 8, 30% PEG 8K	52	0.1 M Citrate pH 6, 0.5 M AmS, 20% PEG 8K
25	0.02 M Tris pH 8, 30% PEG 20K	53	0.2 M Tris pH 7.5, 1 M NaCl, 20% MPD
26	0.75 M NaCit pH 6, 0.4 M NaCl	54	0.2 M Tris pH 7.5, 1 M NaCl, 60% MPD
27	0.03 M MES pH 6, 1 M $MgSO_4$, 20% PEG 8K	55	0.2 M MES pH 6, 0.5 M AmS, 20% MPD
28	0.03 M MES pH 6, 0.1 M $MgSO_4$, 20% PEG 8K	56	0.2 M MES pH 6, 0.5 M AmS, 60% MPD

REFERENCES

Alfano, C. and McMacken, R. (1989a) Ordered assembly of nucleoprotein structures at the bacteriophage λ replication origin during the initiation of DNA replication. *J. Biol. Chem.* **264**:10699-10708

Alfano, C. and McMacken, R. (1989b) Heat shock protein-mediated disassembly of nucleoprotein structures is required for the initiation of bacteriophage λ DNA replication. *J. Biol. Chem.* **264**:10709-10718

Allen, G.C., Jr., Dixon, N.E. and Kornberg, A. (1993) Strand switching of a replicative DNA helicase promoted by the *E. coli* primosome. *Cell* **74**:713-722

Allen, G.C., Jr. and Kornberg, A. (1991) Fine balance in the regulation of DnaB helicase by DnaC protein in replication in *Escherichia coli*. *J. Biol. Chem.* **266**:22096-22101

Allen, G.C., Jr. and Kornberg, A. (1993) Assembly of the primosome of DNA replication in *Escherichia coli*. *J. Biol. Chem.* **268**:19204-19209

Arai, K. and Kornberg, A. (1979) A general priming system employing only *dnaB* protein and primase for DNA replication. *Proc. Natl. Acad. Sci. USA* **76**:4308-4312

Arai, K. and Kornberg, A. (1981a) Mechanism of DnaB protein action. II. ATP hydrolysis by DnaB protein dependent on single- or double-stranded DNA. *J. Biol. Chem.* **256**:5253-5259

Arai, K. and Kornberg, A. (1981b) Mechanism of DnaB protein action. III. Allosteric roles of ATP in the alteration of DNA structure by DnaB protein in priming replication. *J. Biol. Chem.* **256**:5260-5266

Arai, K. and Kornberg, A. (1981c) Mechanism of DnaB protein action. IV. General priming of DNA replication by DnaB protein and primase compared with RNA polymerase. *J. Biol. Chem.* **256**:5267-5272

Arai, K. and Kornberg, A. (1981d) Unique primed start of phage ϕX174 DNA replication and mobility of the primosome in a direction opposite chain synthesis. *Proc. Natl. Acad. Sci. USA* **78**:69-73

Arai, N. and Kornberg, A. (1981e) Rep protein as a helicase in an active, isolatable replication fork of duplex ϕX174 DNA. *J. Biol. Chem.* **256**:5294-5298

Asada, K., Sugimoto, K., Oka, A., Takanami, M. and Hirota, Y. (1982) Structure of replication origin of the *Escherichia coli* K12 chromosome: the presence of spacer sequences in the *ori* region carrying information for autonomous replication. *Nucleic Acids Res.* **10**:3745-3754

Atlung, T., Clausen, E.S. and Hansen, F.G. (1985) Autoregulation of the *dnaA* gene of *Escherichia coli* K12. *Mol. Gen. Genet.* **200**:442-450

Axelrod, N. (1976) Transcription of bacteriophage ϕX174 *in vitro*: Selective initiation with oligonucleotides. *J. Mol. Biol.* **108**:753-770

Baas, P.D., Teertstra, W.R., van der Ende, A. and Jansz, H.S. (1980) Bacteriophage ϕX174 and G4 RF DNA replicative intermediates. A comparitive study using different isolation procedures. *J. Mol. Biol.* **137**:283-299

Baker, T.A. (1995) Replication arrest. *Cell* **80**:521-524

Baker, T.A., Funnell, B.E. and Kornberg, A. (1987) Helicase action of DnaB protein during replication from the *Escherichia coli* chromosomal origin *in vitro*. *J. Biol. Chem.* **262**:6877-6885

Baker, T.A., Sekimizu, K., Funnell, B.E. and Kornberg, A. (1986) Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the *Escherichia coli* chromosome. *Cell* **45**:53-64

Baker, T.A. and Kornberg, A. (1988) Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near *oriC*. *Cell* **55**:113-123

Baker, T.A. and Wickner, S.H. (1992) Genetics and enzymology of DNA replication in *Escherichia coli*. *Annu. Rev. Genet.* **26**:447-477

Biswas, S.B. and Biswas, E.E. (1987) Regulation of DnaB function in DNA replication in *Escherichia coli* by *dnaC* and λ*P* gene products. *J. Biol. Chem.* **262**:7831-7838

Bittner, M. and Vapnek, D. (1981) Versatile cloning vectors derived from the runaway-replication plasmid pKN402. *Gene* **15**:319-329

Bird, R.E., Louarn, J.M., Martuscelli, J. and Caro, L. (1972) Origin and sequence of chromosome replication in *Escherichia coli*. *J. Mol. Biol.* **70**:549-566

Blanar, M.A., Sandler, S.J., Armengod, M-E., Ream, L.W. and Clark, A.J. (1984) Molecular analysis of the *recF* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **81**:4622-4626

Bouché, J.P., Rowen, L. and Kornberg, A. (1978) The RNA primer synthetized by primase to initiate phage G4 DNA replication. *J. Biol. Chem.* **253**:765-769

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248-254

Bramhill, D. and Kornberg, A. (1988a) Duplex opening by DnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* **52**:743-755

Bramhill, D. and Kornberg, A. (1988b) A model for initiation at origins of DNA replication. *Cell* **54**:915-918

Braun, R.E., O'Day, K. and Wright, A. (1985) Autoregulation of the DNA replication gene *dnaA* in *E. coli* K-12. *Cell* **40**:159-169

Braun, R.E. and Wright, A. (1986) DNA methylation differentially enhances the expression of one of the two *E. coli dnaA* promoters *in vivo* and *in vitro*. *Mol. Gen. Genet.* **202**:246-250

Brown, D.R., Roth, M.J., Reinberg, D. and Hurwitz, J. (1984) Analysis of bacteriophage ϕX174 gene A protein-mediated termination and reinitiation of ϕX DNA synthesis. I. Characterisation of the termination and reinitiation reactions. *J. Biol. Chem.* **259**:10545-10555

Brutlag, D., Schekman, R. and Kornberg, A. (1971) A possible role for RNA polymerase in the initiation of M13 DNA synthesis. *Proc. Natl. Acad. Sci. USA* **68**:2826-2829

Campbell, J.L. and Kleckner, N. (1990) *E. coli oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. *Cell* **62**:967-979

Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. (1988) The pMTL*nic*$^-$ cloning vectors. 1. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68**:139-149

Chiaramello, A.E. and Zyskind, J.W. (1989) Expression of *Escherichia coli dnaA* and *mioC* genes as a function of growth rate. *J. Bacteriol.* **171**:4272-4280

Chiaramello, A.E. and Zyskind, J.W. (1990) Coupling of DNA replication to growth rate in *Escherichia coli*: a possible role for guanosine tetraphosphate. *J. Bacteriol.* **172**:2013-2019

Chase, J.W. and Williams, K.R. (1986) Single-stranded DNA binding proteins required for DNA replication. *Annu. Rev. Biochem.* **55**:103-136

Cooper, S. and Helmstetter, C.E. (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* **31**:519-540

Creghton, T.E. (1986) Folding of proteins adsorbed reversibly to ion-exchange resins. *Protein Struct. Folding Des. Proc. GENEX-UCLA Symp.* **39**:249-257

Crooke, E., Thresher, R., Hwang, D.S., Griffith, J. and Kornberg, A. (1993) Replicatively active complexes of DnaA protein and the *Escherichia coli* chromosomal origin observed in the electron microscope. *J. Mol. Biol.* **233**:16-24

Davis, R.W., Botstein, D and Roth, J.R. (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

De Massy, B., Bejar, S., Louarn, J., Louarn, J.M. and Bouché, J.P. (1987) Inhibition of replication forks exiting the terminus region of the *Escherichia coli* chromosome occurs at two loci separated by 5 min. *Proc. Natl. Acad. Sci. USA* **84**:1759-1763

Dodson, M., Roberts, J.D., McMacken, R. and Echols, H. (1985) Specialised nucleoprotein structures at the origin of replication of bacteriophage lambda: complexes with lambda O protein and with lambda O, lambda P, and *Escherichia coli* DnaB proteins. *Proc. Natl. Acad. Sci. USA* **82**:4678-4682

Dodson, M., Echols, H., Wickner, S. Alfano, C., Mensa-Wilmot, K., Gomez, B., LeBowitz, J., Roberts, J.D. and McMacken, R. (1986) Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda: localized unwinding of duplex DNA by a six-protein reaction. *Proc. Natl. Acad. Sci. USA* **83**:7638-7642

Dodson, M., McMacken, R. and Echols, H. (1989) Specialized nucleoprotein structures at the origin of replication of bacteriophage λ. Protein association and disassociation reactions responsible for localized initiation of replication. *J. Biol. Chem.* **264**:10719-10725

Donachie, W.D. (1968) Relationship between cell size and timing of initiation of DNA replication. *Nature* **219**:1077-1079

Dotto, G.P. and Zinder, N.D. (1983) The morphogenic signal of bacteriophage f1. *Virol.* **130**:252-256

Dotto, G.P. and Zinder, N.D. (1984a) Increased intracellular concentration of an initiator protein markedly reduces the minimal sequence required for initiation of DNA synthesis. *Proc. Natl. Acad. Sci. USA* **81**:1336-1340

Dotto, G.P. and Zinder, N.D. (1984b) Reduction of the minimal sequence of initiation of DNA synthesis by qualitative and quantitatine changes of the initiator protein. *Nature* **311**:279-280

Dubeau, L. and Denhardt, D.T. (1981) The mechanism of replication of ϕX174. XVIII. Gene A and A* proteins of ϕX174 bind tightly to ϕX174 replicative form DNA. *Biochim. Biophys. Acta.* **653**:52-60

Dutta, A., Din, S., Brill, S.J. and Stillman, B. (1991) Phosphorylation of replication protein A: A role for cdc2 kinase in G_1/S regulation. *Cold Spring Harbor Symp. Quant. Biol.* **56**:315-324

Dutta, A. and Stillman, B. (1992) cdc2 family kinases phosphorylate a human cell DNA replication factor, RPA, and activate DNA replication. *EMBO J.* **11**:2189-2199

Eisenberg, S. and Ascarelli, R. (1981) The A* protein of ϕX174 is an inhibitor of DNA replication. *Nucleic Acids Res.* **9**:1991-2002

Eisenberg, S. and Finer, M. (1980) Cleavage and circularization of single-stranded DNA: A novel enzymatic activity of ϕX174 A* protein. *Nucleic Acids Res.* **8**:5305-5315

Eisenberg, S. and Kornberg, A. (1979) Purification and characterisation of ϕX174 Gene A protein. A multifunctional enzyme of duplex DNA replication. *J. Biol. Chem.* **254**:5328-5332

Eisenberg, S., Scott, J.F. and Kornberg, A. (1976) Enzymatic replication of viral and complementary strands of duplex DNA of phage ϕX174 proceeds by separate mechanisms. *Proc. Natl. Acad. Sci. USA* **73**:3151-3155

Eisenberg, S., Griffith, J. and Kornberg, A (1977) ϕX174 cistron A protein is a multifunctional enzyme in DNA replication. *Proc. Natl. Acad. Sci. USA* **74**:3198-1202

Eisenberg, S., Scott, J.F. and Kornberg, A. (1978) An enzyme system for replicating the duplex replicative form of ϕX174 DNA. *Cold Spring Harbor Symp. Quant. Biol.* **43**:295-302

Elvin, C.M., Dixon, N.E. and Rosenberg, H. (1986) Molecular cloning of the phosphate (inorganic) transport (*pit*) gene of *Escherichia coli* K12. Identification of the pit^+ gene product and physical mapping of the *pit-gor* region of the chromosome. *Mol. Gen. Genet.* **204**:477-484

Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* **87**:123-126

Erdile, L.F., Collins, K.L., Russo, A., Simancek, P., Small, D., Umbricht, C., Virshup, D., Cheng, L., Randall, S., Weinberg, D., Moarefi, I., Fanning, E. and Kelly, T. (1991) Initiation of SV40 DNA replication: Mechanism and control. *Cold Spring Harbor Symp. Quant. Biol.* **56**:303-313

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **264**:64-71

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1990) Electrospray ionization-principles and practice. *Mass Spectrom. Rev.* **9**:37-70

Ferrige, A.G., Seddon, M.J. and Jarvis, S. (1991) Maximun entropy deconvolution in electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **5**:374-379

Filutowicz, M. and Roll, J. (1990) The requirement of IHF protein for extrachromosomal replication of *Escherichia coli oriC* in a mutant deficient in DNA polymerase I activity. *New Biol.* **2**:818-827

Filutowicz, M., Ross, W., Wild, J. and Gourse, R.L. (1992) Involvement of Fis protein in replication of the *Escherichia coli* chromosome. *J. Bacteriol.* **174**:398-407

Fluit, A.C., Baas, P.D., van Boom, J.H., Veeneman, G.H. and Jansz, H.S. (1984a) Gene A protein cleavage of recombinant plasmids containing the ϕX174 replication origin. *Nucleic Acids Res.* **12**:6443-6454

Fluit, A.C., Baas, P.D., Jansz, H.S., Veeneman, G.H. and van Boom, J.H. (1984b) Gene A protein interacting with recombinant plasmid DNAs containing 25-30 bp of the ϕX174 replication origin. *Adv. Exp. Med. Biol.* **179**:231-240

Fotedar, R. and Roberts, J.M. (1992) Cell cycle regulated phosphoryation of RPA-32 occurs within the replication initiation complex. *EMBO J.* **11**:2177-2187

Fujita, M.Q., Yoshikawa, H. and Ogasawara, N. (1992) Structure of the *dnaA* and DnaA-box region of in the *Mycoplasma capricolum* chromosome: Conservation and variations in the course of evolution. *Gene* **110**:17-23

Fulford, W. and Model, P. (1984) Gene X of bacteriophage f1 is required for phage DNA synthesis. *J. Mol. Biol.* **178**:137-153

Fulford, W. and Model, P. (1988a) Regulation of bacteriophage f1 DNA replication. I. New functions for Gene II and X. *J. Mol. Biol.* **203**:49-62

Fulford, W. and Model, P. (1988b) Bacteriophage f1 DNA replication genes. II. The role of Gene V protein and Gene II protein in complememtary strand synthesis. *J. Mol. Biol.* **203**:39-48

Fuller, R.S. and Kornberg, A. (1983) Purified dnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc. Natl. Acad. Sci. USA* **80**:5817-5821

Fuller, R.S., Funnel, B.E. and Kornberg, A. (1984) The DnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38**:183-190

Funk, F.D. and Snover, D. (1976) Pleiotropic effects of mutants in gene A of bacteriophage ϕX174. *J. Virol.* **18**:141-150

Funnel, B.E., Baker, T.A. and Kornberg, A (1986) Complete enzymatic replication of plasmids containing the origin of *Escherichia coli* chromosome. *J. Biol. Chem.* **261**:5616-5624

Funnel, B.E., Baker, T.A. and Kornberg, A (1987) *In vitro* assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* **262**:10327-10334

Furth, M.E. and Wickner, S.H. (1983) Lambda DNA replication. In Lambda II, ed. R.W. Hendrix, J.W. Roberts, F.W. Stahl, R.A. Weisberg, pp 145-173. New York: Cold Spring Harbor Lab.

Geider, K., Böumel, I. and Meyer, T.F. (1982) Intermediate stages in enzymatic replication of bacteriophage fd duplex DNA. *J. Biol. Chem.* **257**:6488-6493

Gibson, F., Cox, G.B., Downie, J.A. and Radik, J. (1977) A mutation affecting a second component of the F_O portion of the magnesium ion-stimulated adenosine triphosphatase of *Escherichia coli* K12. The *uncC424* allele. *Biochem. J.* **164**:193-198

Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.* **182**:319-326

Gille, H., Egan, J.B., Roth, A. and Messer, W. (1991) The FIS protein binds and bends the origin of chromosomal DNA replication, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* **19**:4167-4172

Gottlieb, P.A., Wu, S., Zhang, X., Teclkenberg, M., Kuempel, P. and Hill, T.M. (1992) Equilibrium, kenetic, and footprinting studies of the Tus-*Ter* protein DNA interaction. *J. Biol. Chem.* **267**:7434-7443

Greenstein, D. and Horiuchi, K. (1987) Interaction between the replication origin and the initiator protein of the filamentous phage f1. *J. Mol. Biol.* **197**:157-174

Greenstein, D., Zinder, N.D. and Horiuchi, K. (1988) Integration host factor interacts with the DNA replication enhancer of filamentous phage f1. *Proc. Natl. Acad. Sci. USA* **85**:6262-6266

Hansen, F.G., Christensen, B.B. and Atlung, T. (1991) The initiator titration model: computer simulation of chromosome and minichromosome control. *Res. Microbiol.* **142**:161-167

Hansen, E.B., Hansen, F.G. and Von Meyenburg, K. (1982a) The nucleotide sequence of the *dnaA* gene and the first part of the *dnaN* gene of *Escherichia coli*. *Nucleic Acids Res.* **10**:7373-7385

Hansen, F.G., Hansen, E.B. and Atlung, T. (1982b) The nucleotide sequence of the *dnaA* gene promoter and of the adjacent *rpmH* gene, coding for the ribosomal protein L34, of *Escherichia coli*. *EMBO J.* **1**:1043-1048

Hansen, F.G. and Rasmussen, K. (1977) Regulation of the *dnaA* product in *Escherichia coli*. *Mol. Gen. Genet.* **155**:219-225

Hansen, F.G. and von Meyenberg, K. (1979) Characterization of the *dnaA*, *gyrB* and other genes in the *dnaA* region of the *Escherichia coli* chromosome on specialized transducing phages λ*tna*. *Mol. Gen. Genet.* **175**:135-144

Hartz, D., McPheeters, D.S. and Gold, L. (1991) Influence of mRNA determinants on translation initiation in *Escherichia coli*. *J. Mol. Biol.* **218**:83-97

Hayashi, M., Fujimura, F.K. and Hayashi, M.N. (1976) Mapping of *in vivo* messenger RNAs for bacteriophage ϕX174. *Proc. Natl. Acad. Sci. USA* **73**:3519-3523

Hayashi, M., Aoyama, A., Richardson, D.L. and Hayashi, M.N. (1988) Biology of the bacteriophage ϕX174, p. 1-71. In R. Calender (ed.), *The Bacteriophages*. Plenum Press, New York.

Heidekamp, F., A.D., Langevald, S.A., Baas, P.D. and Jansz, H.S. (1980) Studies of the recognition sequence of ϕX174 gene A protein. Cleavage site of ϕX gene A protein St-1 RFI DNA. *Nucleic Acids Res.* **8**:2009-2021

Heidekamp, F., A.D., Baas, P.D., van Boom, J.H., Veeneman, G.H., Zipursky, S.L. and Jansz, H.S. (1981) Construction and characterization of recombinant plasmid DNAs containing sequences of the origin of bacteriophage ϕX174 DNA replication. *Nucleic Acids Res.* **9**:3335-3354

Helmstetter, C.E. (1967) Rate of DNA synthesis in the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* **24**:417-427

Helmstetter, C.E. and Cooper, S. (1968) DNA synthesis in the division cycle of rapidly growing *Escherichia coli* B/r. *J. Mol. Biol.* **31**:507-518

Henry, T.J. and Knippers, R. (1974) Isolation and function of the gene A initiator of bacteriophage ϕX174, a highly specific DNA endonuclease. *Proc. Natl. Acad. Sci. USA* **71**:1549-1553

Hiasa, H. and Marians, K.J. (1992) Differential inhibition of the DNA translocation and DNA unwinding activities of DNA helicases by the *Escherichia coli* Tus protein. *J. Biol. Chem.* **267**:11379-11385

Hiasa, H., Sakai, H. and Komano, T. (1989) Identification of single-strand initiation signals in the *terC* region of the *Escherichia coli* chromosome. *FEBS Lett.* **246**:21-24

Hiasa, H., Sakai, H., Komano, T. and Godson, G.N. (1990) Structural features of the priming signal recognised by primase: Mutational analysis of the phage G4 origin of complementary DNA strand synthesis. *Nucleic Acids Res.* **18**:4825-4831

Hidaka, M., Akiyama, M. and Horiuchi, T. (1988) A consensus sequence of the three DNA replication terminus sites on the *E. coli* chromosome is highly homologous to the *terR* sites of the R6K plasmid. *Cell* **55**:467-475

Hidaka, M., Kobayashi, T. and Horiuchi, T. (1991) A newly identified DNA replication terminus site, *TerE*, on the *Escherichia coli* chromosome. *J. Bacteriol.* **173**:391-393

Hill, T.M., Henson, J.M. and Kuempel, P.L. (1987) The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. *Proc. Natl. Acad. Sci. USA* **84**:1754-1758

Hill, T.M., Kopp, B.J. and Kuempel, P.L. (1988) Termination of DNA replication in *Escherichia coli* requires a transacting factor. *J. Bacteriol.* **170**:662-668

Hill, T.M., Teckllenberg, M., Pelletier, A.J. and Kuempel, P.L. (1989) *tus*, the transacting gene required for termination of DNA replication in *Escherichia coli*, encodes a DNA-binding protein. *Proc. Natl. Acad. Sci. USA* **86**:1593-1597

Hochuli, E., Döbeli, H. and Schacher, A. (1987) New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* **411**:177-184

Hirota, Y., Mordoh, J. and Jacob, F. (1970) On the process of cell division in *Escherichia coli*. III Thermosensitive mutants of *Escherichia coli* altered in the process of DNA initiation. *J. Mol. Biol.* **53**:369-387

Horiuchi, K. (1986) Interaction between gene II protein and the DNA replication origin of bacteriophage f1. *J. Mol. Biol.* **188**:215-223

Hughes, P., Landoulsi, A. and Kohiyama, M. (1988) A novel role for cAMP in the control of the activity of the *E. coli* chromosome replication initiator protein, DnaA. *Cell* **55**:343-350

Hwang, D.S. and Kaguni, J.M. (1988) Interaction of DnaA46 protein with a stimulatory protein in replication from the *Escherichia coli* chromosomal origin. *J. Biol. Chem.* **263**:10633-10640

Hwang, D.S. and Kornberg, A. (1992) Opposed actions of regulatory proteins, DnaA and IciA, in opening the replication origin of *Escherichia coli. J. Biol. Chem.* **267**:23087-23091

Ikeda, J., Yudelevich, A. and Hurwitz, J. (1976) Isolation and characterization of the protein coded by *gene A* of bacteriophage ϕX174 DNA. *Proc. Natl. Acad. Sci. USA* **73**:2669-2673

Ikeda, J., Yudelevich, A., Shimamoto, N. and Hurwitz, J. (1979) Role of polymeric forms of the bacteriophage ϕX174 coded Gene A protein in ϕXRFI DNA cleavage. *J. Biol. Chem.* **254**:9416-9428

Itoh, M., Masuda, K., Ito, Y., Akizawa, T., Yoshioka, M., Imai, K., Okada, Y., Sato, H. and Seiki, M. (1996) Purification and refolding of recombinant human prommp-7 (pro-matrilysin) expressed in *Escherichia coli* and its characterisation *J. Biochem.* **119**:667-673

Jacob, F., Brenner, S. and Cuzin, F. (1963) On the regulation of DNA replication in bacteria. *Cold Spring Harbor Symp. Quant. Biol.* **28**:329-248

Janknecht, R., De Martynoff, G., Lou, J., Hipskind, R.A., Nordheim, A. and Stunnenberg, H.G. (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci. USA* **88**:8972-8976

Johnson, S. and Ray, D.S. (1979) Interference between M13 and *ori* M13 plasmids is mediated by a replication enhancer sequence near the viral strand origin. *J. Mol. Biol.* 177:685-700

Kaguni, J.M., Fuller, R.S. and Kornberg, A. (1982) Enzymatic replication of *E. coli* chromosomal origin is bidirectional. *Nature* **296**:623-627

Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal. Chem.* **60**:2299-2301

Katz, L., Kingsbury, D.T. and Helinski, D.R. (1973) Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex. *J. Bacteriol.* **114**:577-591

Khatri, G.S., MacAllister, T., Sista, P.R. and Bastia, D. (1989) The replication terminator protein of *E. coli* is a DNA sequence-specific contra-helicase. *Cell* **59**:667-674.

Klein, A., Lanka, E. and Schuster, H. (1980) Isolation of a complex between the P protein of phage λ and the *dnaB* protein of *Escherichia coli. Eur.J.Biochem.* **105**:1-6

Kobayashi, T., Hidaka, M. and Horiuchi, T. (1989) Evidence of a *ter* specific binding protein essential for the termination reaction of DNA replication in *Escherichia coli. EMBO J.* **8**:2435-2441

Kobori, J.A. and Kornberg, A. (1982) The *Escherichia coli dnaC* gene product. III. Properties of the dnaB-dnaC protein complex. *J. Biol. Chem.* **257**:13770-13775

Kodaira, K., Miyata, T., Suzuki, K., Nakano, K. and Taketo, A. (1989) Possible finger structure in gene A protein of *Microviridae. Biochim. Biophys. Acta.* **1008**:123-124

Kogoma, T. and von Meyenburg, K. (1983) The origin of replication, *oriC*, and the dnaA protein are dispensable in stable DNA replication (*sdrA*) mutants of *Escherichia coli* K-12. *EMBO J.* **2**:463-468

Kogoma, T., Skarstad, K., Boye, E., von Meyenburg, K. and Steen, H.B. (1985) RecA protein acts at the initiation of stable DNA replication in *rnh* mutants of *Escherichia coli* K-12. *J. Bacteriol.* **163**:439-444

Koonin, E.V. (1992) DnaC protein contains a modified ATP-binding motif and belongs to a novel family of ATPases including also DnaA. *Nucleic Acids Res.* **20**:1997

Kornberg, A. (1978) Aspects of DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* 43:1-9

Kornberg, A. and Baker, T.A. (1992) *DNA Replication*, 2nd Edn, W.H. Freeman & Co., New York, N.Y.

Kornberg, A., Baker, T.A., Bertsch, L.L., Bramhill, D., Funnel, B.E., Lasken, R.S., Maki, H., Sekimizu, K. and Wahle, E. (1987) Enzymatic studies of replication of plasmids, p. 137-149. In T.J. Kelly and R. McMacken (ed.), *DNA replication and recombination*. Alan R. Liss, Inc., New York.

Kücherer, C., Lother, H., Kölling, R., Schauzu, M. and Messer, W. (1986) Regulation of transcription of the chromosomal *dnaA* gene of *Escherichia coli*. *Mol. Gen. Genet.* 205:115-121

Kuempel, P.L., Pelletier, A.J. and Hill, T.M. (1989) Tus and the terminators: The arrest of replication in prokaryotes. *Cell* **59**:581-583

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685

Lambert, P.L., Waring, D.A., Wells, R.D. and Reznikoff, W.S. (1986) DNA requirements at the bacteriophage G4 origin of complementary-strand DNA synthesis. *J. Virol.* **58**:450-458

Landoulsi, A., Malki, A., Kern, R., Kohlyama, M. and Hughes, P. (1990) The *E. coli* cell surface specifically prevents the initiation of DNA replication at *oriC* on hemimethylated DNA templates. *Cell* **63**:1053-1060

Langeveld, S.A., van Mansfeld, A.D.M., Baas, P.D., Jansz, H.S., van Arkel, G.A. and Weisbeek, P.J. (1978) Nucleotide sequence of the origin of replication in bacteriophage ϕX174 RF DNA. *Nature* **271**:417-420

Langeveld, S.A., van Mansfeld, A.D.M., J., H.De Winter, J.M. and Weisbeek, P.J. (1979) Cleavage of single-stranded DNA by the A and A* proteins of bacteriophage ϕX174. *Nucleic Acids Res.* **7**:2177-2189

Langeveld, S.A., van Mansfeld, A.D.M., van der Ende, A., van de Pol, J.H., van Arkel, G.A. and Weisbeek, P.J. (1981) The nuclease specificity of the bacteriophage ϕX174 A* protein. *Nucleic Acids Res.* **9**:545-562

LeBowitz, J.H. and McMacken, R. (1986) The *Escherichia coli* DnaB replication protein is a DNA helicase. *J. Biol. Chem.* **261**:4738-4748

LeBowitz, J.H., Zylicz, M., Georgopoulos, C. and McMacken, R. (1985) Initiation of DNA replication on single-stranded DNA templates catalyzed by purified replication proteins of bacteriophage λ and *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **82**:3988-3992

Lee, E.H. and Kornberg, A. (1991) Replication deficiencies in *priA* mutants of *Escherichia coli* lacking the primosomal replication n' protein. *Proc. Natl. Acad. Sci. USA* **88**:3029-3032

Lee, E.H., Kornberg, A., Hidaka, M., Kobayashi, T. and Horiuchi, T. (1989) *Escherichia coli* replication termination protein impedes the action of helicases. *Proc. Natl. Acad. Sci. USA* **86**:9104-9108

Lee, M.S. and Marians, K.J. (1989) The *Escherichia coli* primosome can translocate actively in either direction along a DNA strand. *J. Biol. Chem.* **264**:14531-14542

Legerski, R.J. and Robberson, D.L. (1985) Analysis and optimization of recombinant DNA joining reactions. *J. Mol. Biol.* **181**:297-312

Lehman, I.R. and Kornberg, A. (1992) DNA replication in prokaryotes and eukaryotes. *Chemtracts-Biochem. Mol. Biol.* **3**:1-18

Liberek, K., Georgopoulos, C. and Zylicz, M. (1988) Role of the *Escherichia coli* DnaK and DnaJ heat shock proteins in the initiation of bacteriophage λ DNA replication. *Proc. Natl. Acad. Sci. USA* **85**:6632-6636

Liberek, K., Osipiuk, J., Zylicz, M., Ang, D., Skorko, J. and Georgopoulos, C. (1990) Physical interactions between bacteriophage and *Escherichia coli* proteins required for initiation of lambda DNA replication. *J. Biol. Chem.* **265**:3022-3029

Linney, E. and Hayashi, M. (1973) Two proteins of gene A of ϕX174. *Nature New Biol.* **245**:6-8

Linney, E. and Hayashi, M. (1974) Intergenic regulation of the synthesis of ϕX174 *gene A* cistron. *Nature* **249**:345-348

Løbner-Olesen, A., Atlung, T. and Rasmussen, K.V. (1987) Stability and replication control of *Escherichia coli* minichromosomes. *J. Bacteriol.* **169**:2835-2842

Løbner-Olesen, A., Skarstad, K., Hansen, F.G., Von Meyenburg, K. and Boye, E. (1989) The DnaA protein determines the initiation mass of *Escherichia coli* K-12. *Cell* **57**:881-889

Lother, H., Lurz, R. and Orr, E. (1984) DNA binding and antigenic specifications of DNA gyrase. *Nucleic Acids Res.* **12**:901-914

Louarn, J., Funderburgh, M. and Bird, R.E. (1974) More precise mapping of the replication origin in *Escherichia coli* K-12. *J. Mol. Biol.* **120**:1-5

Love, C.A., Lilley, P.E. and Dixon, N.E. (1996) Stable high-copy-number bacteriophage λ promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* (in press)

Lu, M., Campbell, J.L., Boye, E. and Kleckner, N. (1994) SeqA: a negative modulator of Replication initiation in *E. coli*. *Cell* **77**:413-426

Luria, S.E. and Burrous, J.W. (1957) Hybridization between *Escherichia coli* and *Shigella*. *J. Bacteriol.* **74**:461-476

Maki, H., Maki, S. and Kornberg, A. (1988) DNA polymerase III holoenzyme of *Escherichia coli*. IV. The holoenzyme is an asymmetric dimer with twin active sites. *J. Biol. Chem.* **263**:6570-6578

Malki, A., Hughes, P. and Kohiyama, M. (1991) *In vitro* roles of *Escherichia coli* DnaJ and DnaK heat shock proteins in the replication of *oriC* plasmids. *Mol. Gen. Genet.* **225**:420-426

Mallory, J.B., Alfano, C. and McMacken, R. (1990) Host virus interactions in the initiation of bacteriophage lambda DNA replication. Recruitment of *Escherichia coli* DnaB helicase by lambda P replication protein. *J. Biol. Chem.* **265**:13297-13307

Marians, K.J. (1992) Prokaryotic DNA replication. *Annu. Rev. Biochem.* **61**:673-719

Marians, K.J., Ikeda, J., Schlagman, S. and Hurwitz, J. (1977) *Proc. Natl. Acad. Sci. USA* **74**:1965-1968

Marszalek, J. and Kaguni, J.M. (1994) DnaA protein directs the binding of DnaB protein in initiation of DNA replication in *Escherichia coli*. *J. Biol. Chem.* **269**:4883-4890

Martin, D.F. and Godson, G.N. (1975) Identification of a bacteriophage ϕX174 coded protein involved in the shut-off of host DNA replication. *Biochem. Biophys. Res. Commmun.* **65**:323-330

Masai, H., Bond, M.W. and Arai, K. (1986) Cloning of the *Escherichia coli* gene for primosomal protein i: The relationship to DnaT, essential for chromosomal DNA replication. *Proc. Natl. Acad. Sci. USA* **83**:1256-1260

Masai, H. and Arai, K. (1988) Operon structure of *dnaT* and *dnaC* genes essential for normal and stable DNA replication of *Escherichia coli* chromosome. *J. Biol. Chem.* **263**:15083-15093

Masai, H. and Arai, K. (1989) Leading strand synthesis of R1 plasmid replication *in vitro* is primed by primase alone at a specific site downstream of *oriR*. *J. Biol. Chem.* **264**:8082-8090

Masai, H., Nomura, N. and Arai, K.-I (1990a) The ABC-primosome. A novel priming system employing DnaA, DnaB, DnaC and primase on a hairpin containing a DnaA box sequence. *J. Biol. Chem.* **265**:15134-15144

Masai, H., Nomura, N., Kubota, Y. and Arai, K. (1990b) Roles of ϕX174 type primosome- and G4 type primase-dependent primings in initiation of lagging and leading strand syntheses of DNA replication. *J. Biol. Chem.* **265**:15124-15133

McHenry, C.S. (1988) DNA polymerase III holoenzyme of *Escherichia coli*. *Annu. Rev. Biochem.* **57**:519-550

Meng, C.K., Mann, M. and Fenn, J.B. (1988) Of protons or proteins. *Z. Phys. D - Atoms, Molecules and Clusters* **10**:361-368

Messer, W., Bellekes, U. and Lother, H. (1985) Effect of *dam* methylation on the activity of the *E. coli* replication origin, *oriC*. *EMBO J.* **4**:1327-1332

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) Filamentous coliphage M13 as a cloning vehicle: Insertion of a *Hind*III fragment of the *lac* regulatory region in M13 replicative form *in vitro*. *Proc. Natl. Acad. Sci. USA* **74**:3642-3626

Messer, W., Egan, B., Gille, H., Holz, A., Schaefer, C. and Woelker, B. (1991) The complex of *oriC* DNA with the DnaA initiator protein. *Res. Microbiol.* **142**: 119-125

Meyer, T.F. and Geider, K. (1979a) Bacteriophage fd Gene II-protein. I. Purification, involvement in RF replication, and the expression of Gene II. *J. Biol. Chem.* **254**:12636-12641

Meyer, T.F. and Geider, K. (1979b) Bacteriophage fd Gene II-protein. II. Specific cleavage and relaxation of supercoiled RF from filamentous phages. *J. Biol. Chem.* **254**:12642-12646

Meyer, T.F. and Geider, K. (1982) Enzymatic synthesis of bacteriophage fd viral DNA. *Nature* **296**:828-832

Meyer, T.F., Geider, K., Kurz, C. and Schaller, H. (1979) Cleavage site of bacteriophage fd gene II-protein in the origin of viral strand replication. *Nature* **278**:365-367

Meyer, R.R., Shlomai, J., Kobori, J., Bates, D., Rowen, L., McMacken, R., Ueda, K. and Kornberg, A. (1978) Enzymatic conversion of single-stranded ϕX174 and G4 circles to duplex forms: discontinuous replication. *Cold Spring Harbour Symp. Quant. Biol.* **43**:289-293

Miki, T., Hiraga, S., Nagata, T. and Yura, T. (1978) Bacteriophage λ carrying the *Escherichia coli* chromosomal region of the replication origin. *Proc. Natl. Acad. Sci. USA* **75**:5099-5103

Miki, T., Kimura, M., Hiraga, S., Nagata, T. and Yura, T. (1979) Cloning and physical mapping of the *dnaA* region of the *Escherichia coli* chromosome. *J. Bacteriol.* **140**:817-824

Moffatt, B.A. and Studier, F.W. (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase *Cell* 49:221-227

Mok, M. and Marians, K.J. (1987) The *Escherichia coli* preprimosome and DnaB helicase can form replication forks that move at the same rate. *J. Biol. Chem.* **262**:16644-16654

Monod, J., Cohen-Bazire, G. and Cohn, M. (1951) Sur la biosynthèse de la β-galactosidase (lactase) chez *Escherichia coli*. La specificité de l'induction. *Biochim. Biophys. Acta.* **7**:585-599

Morrison, D.A. (1979) Transformation and preservation of competent bacterial cells by freezing. *Methods Enzymol.* **68**:326-331

Nurse, P., Zavitz, K.H. and Marians, K.J. (1991) Inactivation of the *Escherichia coli* PriA DNA replication protein induces the SOS response. *J. Bacteriol.* **173**:6686-6693

Ogawa, T. and Okazaki, T. (1984) Function of RNaseH in DNA replication revealed by RNaseH defective mutants of *Escherichia coli*. *Mol. Gen. Genet.* **193**:231-237

Ogden, G.B., Pratt, M.J. and Schaechter, M. (1988) The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. *Cell* **54**:127-135

Oka, A., Sugimoto, K., Takanami, M. and Hirota, Y. (1980) Replication origin of *Escherichia coli* K12: the size and structure of the minimum DNA segment carrying information for autonomous replication. *Mol. Gen. Genet.* **178**:9-20

Oka, A., Sugimoto, K., Sasaki, H. and Takanami, M. (1982) An in vitro method generating base substitutions in preselected regions of plasmid DNA: application to structural analysis of the replication origin of the *Escherichia coli* K-12 chromosome. *Gene* **19**:59-69

Oka, A., Sasaki, H., Sugimoto, K. and Takanami, M. (1984) Sequence organisation of replication origin of *Escherichia coli* K-12 chromosome. *J. Mol. Biol.* **176**:443-458

Ollis, D. and White, S. (1990) Protein crystallization. *Methods Enzymol.* **182**:646-662

Pierucci, O, Rickert, M. and Helmstetter, C.E. (1989) DnaA overproduction abolishes cell cycle specificity of DNA replication from *oriC* in *Escherichia coli*. *J. Bacteriol.* **171**:3760-3766

Pelletier, A.J., Hill, T.M. and Kuempel, P.L. (1989) Termination sites T1 and T2 from the *Escherichia coli* chromosome inhibit DNA replication in Col E1-derived plasmids. *J. Bacteriol.* **171**:1739-1731

Polaczek, P. (1990) Bending of the origin of replication of *E. coli* by binding of IHF at a specific site. *New Biol.* **2**:265-271

Polaczek, P. and Wright A. (1990) Regulation of expression of the *dnaA* gene in *Escherichia coli*: Role of the two promoters and the DnaA box. *New Biol.* **2**:574-582

Pritchard, R.H., Barth, P.T. and Collins, J. (1969) Control of DNA synthesis in bacteria. *Symp. Soc. Gen. Microbiol.* **19**:263-297

Ream, L.W., Margossian, L., Clark, A.J. and Von Meyenburg, K. (1980) Genetic and physical mapping of *recF* in *Escherichia coli* K-12. *Mol. Gen. Genet.* **180**:115-121

Roecklein, B., Pelletier, A.J. and Kuempel, P.L. (1991) The *tus* gene of *Escherichia coli*: autoregulation, analysis of flanking sequences, and identification of a complementary system in *Salmonella typhimurium*. *Res. Microbiol.* **142**:169-176

Roepstorff, P. and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrometry* **11**:601

Roth, A. and Messer, W. (1995) The DNA binding domain of the initiator protein DnaA. *EMBO J.* **14**:2106-2111

Russel, D.W. and Zinder, N.D. (1987) Hemimethylation prevents DNA replication in *E. coli*. *Cell* **50**:1071-1079

Samit, C.E., Nansen, F.G., Miller, J.F. and Schaechter, M. (1989) *In vivo* studies of DnaA binding to the origin of replication of *Escherichia coli*. *EMBO J.* **8**:989-993

Sakikabara, Y., Tsukano, H. and Sako, T. (1981) Organization and transcription of the *dnaA* and *dnaN* genes of *Escherichia coli*. *Gene* **13**:47-55

Sakikabara, Y. and Yuasa, S. (1992) Continuous synthesis of the *dnaA* gene product of *Escherichia coli* in the cell cycle. *Mol. Gen. Genet.* **186**:87-94

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual. 2nd Edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5463-5467

Sanhueza, S. and Eisenberg, S. (1985) Bacteriophage ϕX174 A protein cleaves single-stranded DNA and binds to it covalently through a tyrosyl-dAMP phosphodiester bond. *J. Virol.* **53**:695-697

Schaefer, C. and Messer, W. (1991) DnaA protein/DNA recognition. Modulation of the recognition sequence. *Mol. Gen. Genet.* **226**:34-40

Schekman, R., Weiner, J.H., Weiner, A. and Kornberg, A. (1975) Ten proteins required for conversion of ϕX174 single-stranded DNA to duplex form *in vitro*. *J. Biol. Chem.* **250**:5859:5865

Schmitt, J., Hess, H. and Stunnenberg, H.G. (1993) Affinity purification of histidine-tagged proteins. *Mol. Biol. Rep.* **18**:223-230

Scott, J.F., Eisenberg, S., Bertsch, L.L. and Kornberg, A. (1977) Mechanism of duplex DNA replication revealed by enzymatic studies of phage ϕX174: Catalytic strand separation in advance of replication. *Proc. Natl. Acad. Sci. U.S.A.* **74**:193-197

Scovgaard, O. and Hansen, F.G. (1987) Comparison of *dnaA* nucleotide sequences of *Escherichia coli*, *Salmonella typhimurium*, and *Serratia marcescens*. *J. Bacteriol.* **169**:3976-3981

Sekimizu, K., Bramhill, D. and Kornberg, A. (1987) ATP activates DnaA protein in initiation replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* **50**:259-265

Sekimizu, K. and Kornberg, A. (1988) Cardiolipin activation of DnaA protein, the initiation protein of replication in *Escherichia coli*. *J. Biol. Chem.* **263**:7131-7135

Sekimizu, K., Yung, B.Y.-M. and Kornberg, A. (1988) The DnaA protein of *Escherichia coli*. Abundance, improved purification and membrane binding. *J. Biol. Chem.* **263**:7136-7140

Sharma, B. and Hill, T.M. (1992) *TerF*, the sixth identified replication arrest site in *Escherichia coli*, is located within the *rcsC* gene. *J. Bacteriol.* **174**:7854-7858

Shaw, D.C. (1996) John Curtin School of Medical Research, ANU, Canberra. Personal communication

Shepard, H.M., Yelverton, E. and Goeddel, D.V. (1982) Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in the ribosome binding sites. *DNA* **1**:125-131

Shlomai, J. and Kornberg, A. (1980) An *Escherichia coli* replication protein that recognizes a unique sequence within a hairpin region in ϕX174 DNA. *Proc. Natl. Acad. Sci. USA* **77**:799-803

Silhavy, T.J., Berman, M.L. and Enquest, L.W. (1984) *Experiments with Gene Fusions*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Singer, B.S., Gold, L., Shinedling, S.T., Hunter, L.R., Pribnow, D. and Nelson, M.A. (1981) Analysis *in vivo* of translation mutants of the rIIB cistron of bacteriophage T4. *J. Mol. Biol.* **149**:405-432

Sinha, N.K. and Light, A. (1975) Refolding of reduced, denatured trypsinogen and trypsin immobilised on agarose beads. *J. Biol. Chem.* **250**:8624-8629

Skarstad, K., Baker, T.A. and Kornberg, A. (1990) Strand separation required for initiation of replication at the chromosomal origin of *E. coli* is facilitated by a distant RNA-DNA hybrid. *EMBO J.* **9**:2341-2348

Skarstad, K. and Boye, E. (1994) The initiator protein DnaA: evolution properties and function. *Biochim. Biophys. Acta* **1217**:111-130

Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., and Kleckner, N. (1995) *E. coli* SeqA proteinbinds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell* **82**:927-936

Smith, D.W., Garland, A.M., Herman, G., Enns, R.E., Baker, T.A. and Zyskind, J.W. (1985) Importance of state of methylation of *oriC* GATC sites in initiation of DNA replication in *Escherichia coli*. *EMBO J.* **4**:1319-1326

Smith, L.H. and Sinsheimer, R.L. (1976a) The *in vitro* transcrption units of bacteriophage ϕX174. I Characterization of synthetic parameters and measurement of transcrpt molecular weights. *J. Mol. Biol.* **103**:681-697

Smith, L.H. and Sinsheimer, R.L. (1976b) The *in vitro* transcrption units of bacteriophage ϕX174. II *In vitro* initiation sites of ϕX174 transcription. *J. Mol. Biol.* **103**:699-710

Smith, L.H. and Sinsheimer, R.L. (1976c) The *in vitro* transcrption units of bacteriophage ϕX174. III Initiation with specific 5′ end oligonucleotides of *in vitro* ϕX174 RNA. *J. Mol. Biol.* **103**:681-697

Sompayrac, L. and Maaløe, O. (1973) Autorepressor model for control of DNA replication. *Nature New Biol.* **241**:133-135

Stamford, N.P.J., Lilley, P.E. and Dixon, N.E. (1992) Enriched sources of *Escherichia coli* replication proteins. The *dnaG* primase is a zinc metalloprotein. *Biochim. Biophy. Acta.* **1132**:17-25

Studier, F.W. and Moffart, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**:113-130

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**:60-89

Stuitje, A.R., Weisbeek, P.J. and Meijer, M. (1984) Initiation signals for complementary strand synthesis in the region of the replication origin of the *Escherichia coli* chromosome. *Nucleic Acids Res.* **12**:3321-3332

Tabor, S. and Richardson, C.L. (1987) DNA sequence analysis with modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **84**:4767-4771

Theisen, P.W., Grimwade, J.E., Leonard, A.C., Bogan, J.A. and Helmstetter, C.E. (1993) Correlation of gene transcription with the time of initiation of chromosome replication in *Escherichia coli*. *Mol. Microbiol.* **10**:575-584

Tesfa-Selase, F. and Drabble, W.T. (1992) Regulation of the *gua* operon of *Escherichia coli* by the DnaA protein. *Mol. Gen. Genet.* **231**:256-264

Tougu, K., Peng, H. and Marians, K.J. (1994) Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork. *J. Biol. Chem.* **269**:4675-4682

Tsurimoto, T. and Matsubara, K. (1981) Interaction of bacteriophage lambda O protein with the lambda origin sequence. *ICN-UCLA Symp. Mol. Cell. Biol.* **22**:263-275

Tuan, T.L., Cheung, D.T., Wu, L.T., Yee, A., Gabriel, S., Han, B., Morton, L., Nimni, M.E. and Hall, F.L. (1996) Engineering, expression and renaturation of targeted TGF-beta fusion proteins. *Connect. Tissue Res.* **34**:1-9

Van den berg, E.A., Geerse, R.H., Memelink, J., Bovenberg, R.A.L., Magnee, F.A. and Van de Putte, P. (1985) Analysis of regulatory sequences upstream of the *E. coli uvrB* gene; involvement of the DnaA protein. *Nucleic Acids Res.* **13**:1829-1840

van der Avoort, H.G.A.M., Teerststra, R., Versteeg, R. and Weisbeek, P.J. (1983) Genes and regulatory sequences of bacteriophage ϕX174. *Biochim. Biophys. Acta.* **741**:94-102

van der Ende, A., Baker, T.A., Ogawa, T. and Kornberg, A. (1985) Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: primase as the sole priming enzyme. *Proc. Natl. Acad. Sci.* **82**:3954-3958

van der Ende, A., Langevald, S.A., Teerststra, R., van Arkel, G.A. and Weisbeek, P.J. (1981) Enzymatic properties of the bacteriophage ϕX174 A protein on superhelical ϕX174 DNA: A model for the termination of rolling circle DNA replication. *Nucleic Acids Res.* **9**:2037-2055

van Mansfeld, A.D., Langevald, S.A., Weisbeek, P.J., Baas, P.D., van Arkel, G.A. and Jansz, H.S. (1978) Cleavage site of ϕX174 gene A protein in ϕX and G4 RFI DNA. *Cold Spring Harbor Symp. Quant. Biol.* **43**:331-334

van Mansfeld, A.D., Langevald, S.A., Baas, P.D., Jansz, H.S., svan der Marel, G.A., Veeneman, G.H. and van Boom, J.H. (1980) Recognition sequence of bacteriophage ϕX174 gene A protein-the initiator of DNA replication. *Nature* **288**:561-566

van Mansfeld, A.D., van Teeffelen, H.A.A.M., Baas, P.D., Veeneman, G.H., van Boom, J.H. and Jansz, H.S. (1984) The bond in the bacteriophage ϕX174 gene A protein-DNA complex is a tyrosyl-5′-phosphate ester. *FEBS Lett.* **173**:351-356

van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G. (1980) Nucleotide sequence of the filamentous bacteriophage M13 genome: comparison wiht phage fd. *Gene* **11**:129-148

von Meyenburg, K., Hansen, F.G., Nielsen, L.D. and Jørgensen, P. (1977) Origin of replication, *oriC*, of the *Escherichia coli* chromosome: mapping of genes relative to EcoR1 cleavage sites in the *oriC* region. *Mol. Gen. Genet.* **158**:101-109

Wahle, E., Lasken, R.S. and Kornberg, A. (1989a) The dnaB-dnaC replication protein complex of *Escherichia coli*. I. Formation and properties. *J. Biol. Chem.* **264**:2463-2468

Wahle, E., Lasken, R.S. and Kornberg, A. (1989b) The DnaB-DnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. *J. Biol. Chem.* **264**:2469-2475

Wang, Q. and Kaguni, J.M. (1987) Transcriptional repression of the *dnaA* gene of *Escherichia coli* by DnaA protein. *Mol. Gen. Genet.* **209**:518-525

Weiner, J.H., McMacken, R. and Kornberg, A. (1976) Isolation of an intermediate which precedes DnaG RNA polymerase participation in enzymatic replication of bacteriophage ϕX174 DNA. *Proc. Natl. Acad. Sci. USA* **73**:752-756

Wickner, S. and Hurwitz, J. (1974) Conversion of ϕX174 viral DNA to double-stranded chromatography form by purified *Escherichia coli* proteins. *Proc. Natl. Acad. Sci. USA* **71**:4120-4124

Wickner, S. and Hurwitz, J. (1975) Interaction of *Escherichia coli dnaB* and *dnaC*(*D*) gene products *in vitro*. *Proc. Natl. Acad. Sci. USA* **72**:921-925

Wickner, S. and Zahn, K. (1986) Characterisation of the DNA binding domain of bacteriophage λ O protein *J. Biol. Chem.* **261**:7537-7543

Wobbe, C.R., Weissbach, L., Boroweic, J.A., Dean, F.B., Murakami, Y., Bullock, P. and Hurwitz, J. (1987) Replication of SV40-origin-containing DNA with purified proteins. *Proc. Nat. Acad. Sci. USA* **84**:1834-1838

Woelker, B. and Messer, W. (1993) The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* **21**:5025-5033

Wood, C.R., Boss, M.A., Patel, T.P. and Emtage, J.S. (1984) The influence of messenger RNA secondary structure on expression of an immunoglobulin heavy chain in *Escherichia coli*. *Nucleic Acids Res.* **12**:3937-3950

Wu, C.A., Zechner, E.L. and Marians, K.J. (1992a) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. I. Multiple effectors act to modulate Okazaki fragment size. *J. Biol. Chem.* **267**:4030-4044

Wu, C.A., Zechner, E.L., Reems, J.A., McHenry, C.S. and Marians, K.J. (1992b) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. V. Primase action regulates the cycle of Okazaki fragment synthesis by polymerase III holoenzyme. *J. Biol. Chem.* **267**:4074-4083

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and M13mp19 vectors. *Gene* **33**:103-119

Yen, T.S.B. and Webster, R.E. (1981) Bacteriophage f1 gene II and gene X proteins. Isolation and characterisation of the products of two overlapping genes. *J. Biol. Chem.* **256**:11259-11265

Yung, B.Y.-M.and and Kornberg, A. (1989) The DnaA initiator protein binds separate domains in the replication origin of *Escherichia coli*. *J. Biol. Chem.* **264**:6146-6150

Yung, B.Y.-M., Crooke, E. and Kornberg, A. (1990) Fate of the initiator protein in replication at the origin of the *E. coli* chromosome *in vitro*. *J. Biol. Chem.* **265**:11622-11627

Yuasa, S. and Sakikabara, Y. (1980) Identification of the *dnaA* and *dnaN* gene products of *Escherichia coli*. *Mol. Gen. Genet.* **180**:267-273

Zahn, K. and Blattner, F.R. (1985) Binding and bending of the λ replication origin by the phage O protein *EMBO J.* **4**:3605-3616

Zechner, E.L., Wu, C.A. and Marians, K.J. (1992a) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. *J. Biol. Chem.* **267**:4045-4053

Zechner, E.L., Wu, C.A. and Marians, K.J. (1992b) Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. III. A polymerase-primase interaction governs primer size. *J. Biol. Chem.* **267**:4054-4063

Zinder, N.Z. and Horiuchi, K. (1985) Multiregulatory element of filamentous bacteriophages. *Micobiol. Rev.* **49**:101-106

Zylicz, M., LeBowitz, J.H., McMacken, R. and Georgopoulos, C. (1983) The DnaK protein of *Escherichia coli* possesses an ATPase and autophosphorylating activity and is essential in an *in vitro* DNA replication system. *Proc. Natl. Acad. Sci. USA* **80**:6431-6435

Zyskind, J.W., Cleary, J.M., Brusilow, W.S., Harding, N.E. and Smith, D.W. (1983) Chromosomal replication origin from the marine bacterium *Vibrio harveyi* functions in *Escherichia coli: oriC* consensus sequence. *Proc. Natl. Acad. Sci. USA* **80**:1164-1168